STRUCTURE AND FUNCTION PREDICTION OF HUMAN MUSCARINIC ACETYLCHOLINE RECEPTOR 1, CATION-π STUDIES, AND PROTEIN DESIGN

Thesis by
Joyce Yao-chun Peng

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2005

(Defended May 17, 2005)

ACKNOWLEDGEMENTS

First of all, I want to thank my advisors Bill Goddard and Vaidehi Nagarajan for their continuous advice, support, and guidance in my scientific and professional development since 1997. They have shaped me and I have learned so much from them!

I also want to thank Bill Goddard and Ray Deshaies for giving me the opportunity to come back to Caltech to finish my PhD study. I want to thank Mary Kennedy, Marianne Bronner-Fraser, and Judith Campbell for taking the time to participate in my PhD thesis committee and for their valuable advice. The following members in the Goddard group have been instrumental to my research: Dee Datta, Spencer Hall, Huazhang Shen, Rene Trabanino, Peter Freddolino, Chris McClendon, Julius Su, Victor Kam, Yashar Kalani, Youyong Li, Fangqiang Zhu, Jiyoung Heo, Sam Cheung, Weiqiao Deng, Peter Huskey, Frank Ducheneaux, and Shantanu Sharma. I want to thank my parents for encouraging me to come back to Caltech and my boyfriend, Wilfred Lee, for his continuous support when I encountered research challenges.

ABSTRACT

Muscarinic acetylcholine receptors, a G protein-coupled receptor, are responsible for a wide range of diseases. We predicted the 3D structure of the human M1 muscarinic receptor using the MembStruk method and validated its binding sites for 10 agonists and antagonists using the HierDock method. The predicted binding sites, the intramolecular contacts that stabilize the receptor conformation, and the *in silico* mutagenesis results, agree well with mutagenesis data. The calculated relative binding energies correlate well with measured binding affinities. In addition, the predicted binding sites provide a structural basis for the large reduction in ligand binding affinity and signaling efficacy by Trp 157 and Pro 159 mutations, which was not previously explained by homology models. The predicted binding sites illustrate the importance of aromatic residues in ligand binding through extensive cation-π and aromatic-aromatic interactions, with new mutation candidates suggested. The predicted M1 structure improves our understanding of the muscarinic receptors, offers a basis for structure based drug design, and is a successful step toward applying these procedures in predicting the structures of other muscarinic receptor subtypes.

We used high-level quantum mechanics to quantify cation-π interactions in the crystal structure of carbamylcholine binding to Acetylcholine-binding Protein, a nicotinic receptor homolog. The calculated effects of fluorinated unnatural amino acid substitutions also correlate excellently with experimental EC50 data, suggesting that quantum mechanics

can accurately predict cation-π binding in a protein environment and provides a good model system in developing force fields to better describe cation-π interactions.

Histidines are known to modulate pH responsive binding. We designed a series of histidine derivatives by substituting its imidazole ring with functional groups that are small in size and lack the ability to form hydrogen bonds. Quantum mechanical calculations of the acid dissociation constants (pKa) show that these substitutions shift the histidine pKa upward or downward. We report a list of histidine derivatives and their corresponding pKa values that can be used in designing tumor specific drugs (e.g. HER2-Herceptin antibody), drug delivery through pH sensitive hydrogels, drug recycling, catalysis, and biosensors development. An example of how these unnatural histidines can be used is illustrated with 2-methyl histidine incorporated in a c-Myc-Max heterodimer.

TABLE OF CONTENTS

Part 1

Structure and Function Prediction of Human Muscarinic Acetylcholine Receptor 1 (M1), a G Protein Coupled Receptor

Abstract

The five subtypes of muscarinic acetylcholine receptors are implicated in the pathogenesis of a wide range of diseases including cognitive dysfunction and smooth muscle disorder. Availability of the 3D structure of these receptors would greatly enhance subtype specific drug design. In this study, we predicted the structure of the human muscarinic acetylcholine receptor 1 (M1) using the MembStruk computational method and validated this structure using the HierDock docking method to predict the ligand-binding sites for 10 agonists and antagonists. The predicted binding sites, the intramolecular contacts that stabilize the receptor conformation, and the results of *in silico* mutagenesis agree well with experimental mutagenesis data. The calculated relative binding energies also correlate well with measured binding affinities. The predicted binding site of antagonists is located between TM helices 3, 4, 5, 6, and 7, whereas the agonist binding site involves residues from helices 3, 6, and 7. In addition, the antagonist binding sites provide a structural basis for the large reduction in ligand binding affinity and signaling efficacy by mutations of Trp 157 [4] and Pro 159 [4], which were not previously explained by other homology based models of M1. The numbers shown in the square parentheses are the TM helix numbers. Our prediction suggests that Trp 157, directly participates in ligand binding, instead of forming a secondary ligand docking site, while Pro 159 provides an indirect conformational switch to position Trp 157. In addition to the critical Asp 105 [3], several aromatic residues including Tyr 381 [6], Tyr 404 [7], Tyr 408 [7] are important in binding the ligand quaternary ammonium head group through cation-π interactions with a geometry

similar to that in the nicotinic acetylcholine receptor crystal structure complex. Our results also suggest that upon binding of ligands with a charged tertiary amine head group, proton transfer from the ligand to Asp 105 [3] might occur. In addition, an extensive aromatic network involving Tyr 106 [3], Trp 157 [4], Phe 197 [5], Trp 378 [6], and Tyr 381 [6] is important in stabilizing the antagonists binding sites. For antagonists with two terminal aromatic rings, this aromatic network extends to Trp 164 [4], Tyr 179 [extracellular loop 2], and possibly Phe 391 [6]. Asn 382 [6] forms hydrogen bonds with selected antagonists. Tyr 381 [6] and Ser 109 [3] form hydrogen bonds with the ester moiety of acetylcholine, which binds in the gauche conformation. Many residues whose Ala mutations were shown to decrease the receptor stability and/or induce agonist-independent basal signaling activity also form favorable intramolecular contacts in the predicted structure. This study provides the structural insights for improved understanding of the M1 receptor and its binding sites and is a successful step toward applying these procedures to predict the structures and function of the other four muscarinic receptor subtypes, which should, in turn, enable the development of subtype-specific agonists and antagonists with reduced side effects.

Chapter 1

INTRODUCTION

The pharmacological action of acetylcholine is divided into two main categories: muscarinic and nicotinic (1). The nicotinic acetylcholine receptors belong to the family of ligand-gated ion channels. The muscarinic acetylcholine receptors, which belong to the G protein-coupled receptor (GPCR) family, play a critical role in mediating the metabotropic action of acetylcholine. Among the five subtypes of muscarinic receptors, the M1, M3, and M5 receptors are stimulatory and couple preferentially to G-proteins of the Gq/G11 class, which leads to phosphoinositide breakdown. The M2 and M4 receptors are inhibitory and couple primarily to G-proteins of the Gi and Go classes, leading to adenylate cyclase inhibition that reduces the cytoplasmic concentrations of cyclic AMP (1). The M1-M5 receptor knockout mice study suggested that the M1 receptor modulates neurotransmitter signaling in the cortex and hippocampus. The M2 receptor mediates tremor, hypothermia, and auto-inhibition of acetylcholine release in several brain regions. The M3 receptor is involved in exocrine gland secretion, smooth muscle contractility, and weight gain. The M4 receptor modulates dopamine activity in motor tracts and acts as an inhibitory auto-receptor in the striatum. The M5 receptor modulates central dopamine function and the tone of cerebral blood vessels (2).

The muscarinic receptors are responsible for a wide range of diseases, including cognitive dysfunctions such as Alzheimer's disease (3), schizophrenia, and Parkinson's disease (2), as well as smooth muscle disorders such as overactive bladder, irritable bowel syndrome, and chronic obstructive pulmonary disease (1). Significant effort has been focused on identifying M1 agonists and M2 antagonists of muscarinic acetylcholine receptors as a potential treatment of Alzheimer's disease by stimulating the cholinergic system (3, 4). The rationale for developing M1 agonists was based on the role of acetylcholine in learning and memory function and the finding that cholinergic neurons degenerated in Alzheimer's patients. In addition, M1 receptors promote α-secretase activity, which results in the secretion of non-toxic peptide fragments of amyloid precursor protein and decreased Aβ deposition (3, 4).

Despite the extensive development of muscarinic M1 receptor agonists for the treatment of cognitive dysfunctions, many drug candidates were discontinued because of the occurrence of undesirable side effects from cross reactivity to other receptor subtypes (3). The high sequence homology between muscarinic receptor subtypes contributes to the relatively poor selectivity of compounds that display muscarinic agonist activity. Despite the importance of muscarinic receptors, no 3D crystal structure is currently available for any muscarinic receptor subtype that could enable structure-based drug design to develop drugs with improved subtype specificity and reduced side effects. To this date the only experimental 3D structure available for GPCR is bovine rhodopsin (5). The lack of crystal structures is because the GPCRs are bound to the membrane, making it difficult to express

them in sufficient quantities for crystallization. In addition, it is difficult to construct reliable homology models of muscarinic receptors because of the low sequence homology (16% overall sequence identity and 21% in the TM regions for M1 receptor) between bovine rhodopsin and muscarinic receptors. Because the binding site of 11cis-retinal in bovine rhodopsin is narrow, in building M1 homology models, manual adjustments of the TM helices or the second extra-cellular (EC2) loop positions were often required to fit the ligands, introducing biases to the model (6,7). Attempts to use the M1 homology model in virtual ligand screening also generated poor hit rates for agonists and antagonists (6). Fortunately, extensive mutagenesis experiments and ligand binding data is available for the muscarinic receptors, especially the M1 receptor (8-12).

We report here the predicted structure of human M1 muscarinic receptor using the Membstruk computational method. MembStruk, unlike homology based models, does not use any information from the crystal structure of bovine rhodopsin. We then used the HierDock method to predict the ligand binding sites and calculated the binding energies of 10 agonists and antagonists to the M1 receptor. The MembStruk and HierDock methods have been validated for bovine rhodopsin (13, 14), human β2 adrenergic receptor (15), human D2 dopamine receptor (16), and mouse and human olfactory receptors (17-20), where the predicted binding sites are in good agreement with the experimental results.

In this study, the predicted M1 structure and the predicted binding sites of agonists and antagonists agree well with experimental results. The calculated relative binding

energies also correlate well with measured binding affinities. Among the 8 agonists and antagonists studied, agonists bind at a region between TM 3, 6, 7, whereas antagonists bind at a region between TM 3, 4, 5, 6, and 7, which overlaps with the agonists binding region. Our results show that Asp 105 [3] is essential in ligand binding. This effect is even more significant for ligands with basic tertiary amine than with quaternary ammonium head groups. The numbers shown within square parentheses are the TM helix to which the residue belongs. Tyr 381 [6], Tyr 404 [7], Tyr 408 [7] are also important in stabilizing the ligand quaternary ammonium head group with cation-π interactions. The cation-π distances at the predicted M1 binding site are very similar to the ones in the nicotinic acetylcholine receptor crystal structure complex (21). Extensive aromatic networks involving Tyr 106 [3], Trp 157 [4], Phe 197 [5], Trp 378 [6], and Tyr 381 [6] are important in stabilizing the antagonist binding sites. For antagonists with two terminal aromatic rings, the aromatic network extends to Trp 164 [4], Tyr 179 [extra-cellular loop 2], and possibly Phe 391 [6]. In addition to these favorable aromatic-aromatic interactions, Asn 382 [6] forms hydrogen bonds to antagonists such as NMS, pirenzepine, and QNB. Tyr 381 [6] and Ser 109 [3] form hydrogen bonds with the ester moiety of acetylcholine, which binds in the *gauche* conformation. In addition, many residues whose Ala mutations were shown to decrease the receptor stability and/or induce agonist-independent basal signaling activity form favorable intra-molecular contacts in the predicted structure.

Chapter 2 of my thesis presents the results of the predicted M1 structure and its binding sites for the 10 agonists and antagonists studied. The results are compared with

available mutagenesis and ligand binding data. Several new mutation candidates are suggested. Possible mechanisms of receptor activation by comparing agonist and antagonist binding sites are discussed. Chapter 3 provides a detailed description of the methods used to predict the M1 structure, with several refinements to the Membstruk method listed. This study not only provides the structural insights for improved understanding of the M1 receptor and its binding sites, but also is a successful step towards applying these procedures to predict the 3D structures and function of the other four subtypes of muscarinic receptors, which should, in turn, enable the development of subtype-specific agonists and antagonists with reduced side effects.

Chapter 2

TERTIARY STRUCTURE AND LIGAND BINDING SITES OF THE HUMAN M1 RECEPTOR: EXPERIMENTAL VALIDATIONS AND NEW INSIGHTS

Results

Properties of the Predicted Human M1 Receptor Structure

The three dimensional structure of M1 is shown in Fig. 1. Among the residues whose Ala substitution (7) reduced the level of expression of M1 receptor binding sites by more than 30 fold to less than 3% of the wild type level (by rendering the receptor unstable), the conserved Trp 150 [4] points towards TM 2 and TM 3 (Fig. 2A) in our predicted structure, similar to its position in the bovine rhodopsin crystal structure. Asp 71 [2], the most conserved Asp across many GPCRs on TM2, is hydrogen bonded to Asn 43 [1] (Fig. 2B). Based on Ala mutagenesis experiments, Asp122 [3] in the conserved DRY sequence was suggested to make intra-molecular contacts whose integrity is important for efficient receptor folding (22). In the predicted structure, Asp 122 [3] forms a salt bridge to both Arg 365 [6] and Arg 210 [5] at the intra-cellular end of the receptor (Fig. 2C). Mutation of Leu 116 [3] not only reduced the receptor expression level but also increased agonist affinity and raised constitutive activity, a phenotype suggested for residues forming intra-molecular contacts that are important in stabilizing the inactive state, but not the

activated state of the receptor (7). In our predicted structure, Leu 116 [3] is in the core of the receptor and contacts Phe 374 [6], which also contacts Asn 414 [7] (Fig. 2D). Mutation of Phe 374 [6] and Asn 414 [7] on the M5 receptor, has been reported to cause a 10 fold decrease in the receptor expression level (7). Mutation of the residue equivalent to M1 Phe 374 [6] in the M5 receptor was also reported to increase constitutive activity (7). The close proximity of Leu 116 [3], Phe 374 [6], and Asn 414 [7] in the predicted structure corresponds well to the experimental results where the His-substitution of these three residues formed a Zn2+ binding site in the ground state of the M1 receptor (Fig. 2D) (23). Note that the α-carbons of these three residues must be separated by less than 13 Å to allow the corresponding imidazole side chains to coordinate a metal ion (23). The predicted α-carbon distances among these residues are 7.7 Å, 8.4 Å, and 11.0 Å thus satisfying this criterion (Fig. 2D). Among other residues that also increased acetylcholine affinity and basal signaling activity (but to a lesser extent), Trp 405 [7] forms an inter-helical hydrogen bond with Thr 32 [1] (Fig. 2E), Ala 160 [4] forms a hydrophobic interaction with Thr 192 [5] (Fig. 2F), residue shown to affect agonist but not antagonist binding (24), and Ser 120 [3] points toward the core of the receptor and contacts both Leu 116 [3] and Phe 374 [6]. Several inter-helical hydrogen bonds are present at the ligand binding sites (Fig. 3) in the predicted structure, including the ones between Asp 105 [3] and Tyr 408 [7], between Ser 109 [3] and Tyr 381 [6], and between Tyr 389 [6] and Tyr 404 [7].

The hydrophobicity profile of the multiple sequence alignment used for predicting the length of the TM regions of the M1 receptor is shown in Fig. 4. The position of

maximum hydrophobicity in the hydrophobicity profile is called the TM hydrophobic center. Fig. 5 lists the TM regions and the hydrophobic center of each helix predicted by Membstruk hydrophobicity profiling and capping procedures (14). The helical bending angles and helical tilt of the predicted M1 structure are listed in Fig. 6.

Predicted Binding Sites for Agonists and Antagonists in Human M1 Receptor

We have predicted the binding sites of 10 agonists and antagonists (shown in Fig. 7) and calculated their binding energies using the HierDock 2.0 method (13).

Scanning the Entire M1 Receptor for Binding Sites

To locate the binding site of the agonists and antagonists, the void region in the entire receptor structure was divided into 27 regions and the HierDock method was used for 5 ligands in each box to determine the binding site. The ligands used included NMS, QNB, acetylcholine, tetramethylammonium, and 4-DAMP. For each ligand, the best scoring five conformations in energy with at least 80% buried surface area were used to select the best regions for subsequent docking of all ligands. The positions of the top 2 best binding regions selected are shown in Fig. 8. All the 8 agonists and antagonists were subsequently docked into this best binding region and the binding energies calculated using HierDock 2.0 method as briefed in the Methods section.

Predicted Binding Site for Antagonist N-methylscopolamine (NMS)

Fig. 9 shows the predicted binding site of the antagonist, the active S-(−)-enantiomer of NMS, determined by HierDock. The predicted NMS binding site involves residues in TM 3 to 7. Figures 9A and 9B show that the NMS binds between TM helices 3, 4, 5, 6 and 7 in the predicted M1 receptor structure. Figure 9C shows the residues within 4 Å of the ligand in the predicted binding site of NMS. The charged quaternary ammonium headgroup forms a 3.5 Å salt bridge to Asp 105 [3]. It also interacts with the surrounding aromatic residues Tyr 404 [7], Tyr 408 [7], Trp 101 [3], and Tyr 381 [6] primarily through cation-π interactions (46). The hydroxylmethyl group on the tropic acid side chain of NMS is hydrogen-bonded to Asn 382 [6] at a distance of 2.9 Å. The carbonyl group of the ester moiety is hydrogen-bonded to Ser 109 [3] at a distance of 2.8 Å. The benzene group at the tropic acid side chain of NMS interacts with the surrounding aromatic residues Tyr 106 [3], Trp 157 [4], Trp 378 [6], and Phe197 [5]. Fig. 9C highlights the favorable aromatic-aromatic interactions between the negatively charged carbon atoms at the NMS benzene ring and the positively charged hydrogen atoms of Tyr 106 [3] and Trp 157 [4] aromatic rings.

Computational Ala Scanning Mutations for NMS Binding Site in the M1 Receptor

We performed systematic mutation of the residues in the binding site of NMS (all residues within 6Å) to Ala and calculated the change in binding energy, to compare to the experimental Ala substitution mutagenesis data on NMS binding affinity. We observed that the Ala mutation of the following residues changes the binding energies by more than 5

kcal/mol: Asp 105 [3], Tyr 106 [3], Tyr 381 [6], Asn 382 [6], Tyr 404 [7], and Tyr 408 [7]. All these residues were shown experimentally to affect NMS binding affinity by more than 30 fold (7).

Predicted Binding Site for NMS Derivatives

Fig. 10 shows the 4 Å binding site of three NMS derivatives shown in Fig. 7: N-Methylatropine, N-Methylhomatropine, and N-Methylacetyltropine. These three derivatives of NMS also bind in the region between TM helices 3, 4, 5, 6 and 7. In N-Methylhomatropine, a methylene group bearing the hydroxyl group (predicted to form a hydrogen bond with Asn 382 [6] in NMS) was removed from N-Methylatropine (Fig. 7). In comparing the binding affinities of N-Methylatropine and N- Methylhomatropine, it was shown experimentally that removing this methylene group reduced binding 90 fold (8). In N-Methylacetyltropine, the hydroxylmethyl and benzene ring were removed and the binding affinity was further reduced (8). We observed a similar trend in our predictions where the calculated relative binding energy (in absolute value) decreased from 28.55 kcal/mol for N-Methylatropine (pIC50=9.56±0.04) to 21.92 kcal/mol for N-Methylhomatropine (pIC50=7.61±0.04), and further to 12.63 kcal/mol for N-Methylacetyltropine (pIC50=5.44±0.04), which correlates well with the change in measured binding affinities for these NMS derivatives. It should be noted that the calculated binding energies do not include the room temperature effects, or explicit entropic terms.

Predicted Binding Site for Antagonist 4-DAMP

The binding site of the antagonist 4-DAMP is located between the TM helices 3, 4, 5, 6, and 7. Fig. 11 shows the 4 Å binding site of 4-DAMP, or 4-diphenylacetoxy-N-methylpiperidine methiodide. The quaternary ammonium head group forms a 3.8Å weak salt bridge with Asp 105 [3], which is expected for a quaternary ammonium group. The surrounding aromatic residues Tyr 404 [7], Tyr 408 [7], and Tyr 381 [6] interact with the ligand head group through cation-π interactions (Fig. 11). The carbonyl group of the ester moiety forms a hydrogen bond to Ser 109 (2.9 Å). Ser 109 also forms an inter-helical hydrogen bond with Tyr 381 (2.9 Å). The two benzene rings of 4-DAMP and surrounding aromatic residues Tyr 106 [3], Trp 157 [4], Trp 378 [6], Phe 197 [5], and Tyr 381 [6] form an aromatic network that stabilizes the 4-DAMP binding site.

Predicted Binding Site for Antagonist Quinuclidinyl Benzilate (QNB)

Fig. 12 shows the residues within the 4 Å binding site of (-)-QNB. The predicted binding site of QNB is located between helices 3, 4, 5, 6, and 7. The charged protonated tertiary amine group forms a salt bridge with Asp 105 [3] at a distance of 2.7 Å. The hydroxyl group of QNB is hydrogen bonded to Asn 382 [6] at a distance of 2.9 Å. The ester moiety of QNB is hydrogen bonded (2.9 Å) to Ser 109 [3]. As in 4-DAMP, the QNB benzene ring is caged by the surrounding aromatic residues Tyr 106 [3], Trp 157 [4], Trp 378 [6], Phe 197 [5], and Tyr 381 [6] that stabilize the binding site.

Predicted Binding Site for Antagonist Pirenzepine

Fig. 13 shows the 4 Å binding site of pirenzepine. The pirenzepine binding site involves residues from TM helices 3, 4, 5, 6, and 7. The charged protonated tertiary amine group of pirenzepine forms a salt bridge (2.9 Å) with Asp 105 [3]. The lactam group of pirenzepine forms a hydrogen bond (2.9 Å) with Asn 382 [6]. As in other ligands, the pirenzepine benzene ring and the surrounding aromatic residues Tyr 106 [3], Trp 157 [4], Phe 197 [5], Trp 378 [6], and Tyr 381 [6] form an aromatic cage that stabilizes the binding site.

Predicted Binding Site for Acetylcholine, the Endogenous Agonist

The predicted binding site of acetylcholine is located between TM helices 3, 6, and 7 as shown in Fig. 14. There is a salt bridge (3.9 Å) between Asp 105 [3] and the quaternary ammonium head group of acetylcholine. The head group is also stabilized by the surrounding aromatic sidechains Tyr 381 [6], Tyr 404 [7], and Tyr 408 [7] through cation-π interactions (Fig. 13). Trp 101 [3] is also in close proximity to the acetylcholine head group. The carbonyl group of the ester moiety of acetylcholine is hydrogen bonded to both Tyr 381 (2.9 Å) and Ser 109 (2.8 Å).

The Effect on Acetylcholine Binding of Alanine Mutations in M1

We performed systematic mutation of the residues in the binding site of acetylcholine (all residues within 6 Å) to Ala and calculated the change in binding energy, to compare to the Ala substitution mutagenesis data on acetylcholine binding affinity. We observed that the following residues changed the binding energies by more than 2.5 kcal/mol: Asp 105

[3], Tyr 106 [3], Ser 109 [3], Tyr 381 [6], Tyr 404 [7], and Tyr 408 [7]. The Ala mutations of these residues significantly reduced acetylcholine binding affinities by 70, 60, 10, 40, 30, and 30 fold, respectively (8, 10, 11).

Predicted Binding Site for Agonist Oxotremorine-M

Fig. 15 shows the 4 Å binding site of oxotremorine-M. This agonist binds in a similar location as acetylcholine between helices 3, 6 and 7. There is a salt bridge (3.9 Å) between Asp 105 [3] and the quaternary ammonium head group of oxotremorine-M. The head group is also stabilized by the surrounding aromatic side chains including Tyr 381 [6], Tyr 404 [7], and Tyr 408 [7] through cation-π interactions (Fig. 14). The carbonyl group of the ester moiety of oxotremorine-M is hydrogen bonded to both Tyr 381 [6] and Ser 109 [3], with a distance of 2.9 Å.

Predicted Binding Site for Agonist Tetramethylammonium

The tetramethylammonium ion is also an agonist (pIC50 = 2.9) and binds similarly to the acetylcholine head group (Fig. 16). The tetramethylammonium ion is stabilized by both a salt bridge (3.6 Å) with Asp 105 [3] and by cation-π interactions with the surrounding aromatic residues Tyr 381 [6], Tyr 404 [7], and Tyr 408 [7]. The distances of the cation-π interactions are listed in Fig. 15.

Nature of Agonist Interaction Compared to the Antagonist Interaction with the M1 Receptor

To understand quantitatively how the agonist interacts with the receptor compared to the antagonist, we have calculated the interaction energy of all the residues in the binding site (within 6Å) with the ligand. The summation of the contribution from each residue on a particular TM helix to the interaction energy shows how strongly the ligand interacts with that particular TM helix. The results for antagonists with a quaternary ammonium head group (NMS and 4-DAMP) are listed in Fig. 17 and Fig. 18. The results for antagonists with the basic tertiary amine head group (QNB and pirenzepine) are listed in Fig. 19 and Fig. 20. The results for acetylcholine are listed in Fig. 21 and Fig. 22. Our results suggested that the interaction energy of the antagonists with TM 3 and TM 6 are strong, followed by TM 7. The Asp 105 [3] salt bridge plays a much more significant role in binding the antagonists with a tertiary amine head group than in binding the ones with a quaternary ammonium head group.

Discussion

Antagonist Binding: Comparison of NMS Binding Site with Mutagenesis Data

The effect of Ala mutations for the residues in TM 2, 3, 4, 5, 6 and 7 on the binding affinity of NMS (Kd=125 pM) has been well studied experimentally (7, 25). The predicted 4 Å NMS binding site includes eight of the nine ligand-anchor residues whose Ala substitution mutation caused >30-fold reduction in NMS affinity in the scanning

mutagenesis experiments. These residues include Asp 105 [3], Tyr 106 [3], Trp 157 [4], Tyr 381 [6], Asn 382 [6], Tyr 404 [7], Cys 407 [7], and Tyr 408 [7] (Fig. 9C). The only exception is Pro 159 [4], whose Ala substitution mutation caused a 63-fold reduction in NMS binding (10), but caused a very small reduction (less than 3%) in calculated binding energy. In the predicted structure, Pro 159 [4] has a secondary effect on NMS binding by inducing a bend in TM4 that helps to position Trp 157 [4] at the NMS binding site. When Pro159 is mutated to Ala computationally, the bend in the backbone caused by proline is kept and the helical structure has not been reoptimized for Ala instead of Pro. This substitution may give rise to a different helical structure for TM4. Hence the calculated effect on the binding energy due to Ala substitution of Pro 159 is small. However, experimentally such a substitution might change the nature of the bend in TM4. Trp 157 [4] was implicated to be important in ligand binding in M1. Its Ala substitution mutation caused a 123 fold reduction in NMS binding (10), a 132 fold reduction in 4-DAMP binding (12), and a 330 fold reduction in QNB binding affinity (10). In the predicted structure we found that Trp157 [4] is directly involved in binding of these antagonists.

Placement of TM 4 in the predicted M1 structure is critical to explaining the experimental observations on Trp157. The predicted M1 structure is different from the rhodopsin-based homology model in the relative translational position of TM4. In the MembStruk predicted M1 structure, Trp 157 [4] is located next to Tyr 106 [3] at the predicted binding site, whereas in homology models, Trp 157 [4] was placed near the N-terminal end of TM3, close to Asp 99 [3]. Since the homology models do not directly

explain the effect of Trp157 mutation experiments, it was postulated that Trp 157 might form a putative secondary docking site before ligands enter the main binding site (7, 10). However, the MembStruk predicted structure suggests that Trp 157 [4], instead of forming a secondary docking site, directly interacts with the antagonists at the central binding site through aromatic-aromatic interactions. Our results provide a structure basis for explaining the strong effects of these mutations on ligand binding. The step of the Membstruk procedure that optimizes the relative translation of the helices by aligning the hydrophobic centers (the position of the maximum hydrophobicity as shown in Fig. 5) for each helix to a plane, is essential in determining the ligand binding sites as shown by Trp157 [4].

Common Features of Antagonist Binding Sites

Antagonists with a Charged Quaternary Ammonium Head Group

Many M1 antagonists, including NMS and 4-DAMP (Ki = 900 pM), have a charged quaternary ammonium head group. In the binding sites of both NMS and 4-DAMP, the positively charged quaternary ammonium head group of the ligand is stabilized by the negatively charged Asp 105 of TM3 and the surrounding tyrosine residues (Tyr 381, Tyr 404, Tyr 408) of TM 6 and 7, primarily though cation-π interactions.

Binding Site of Charged Tertiary Amine Group in Antagonists

Antagonists QNB (Kd=16 pM) and pirenzepine (Ki = 4.4 nM) have a tertiary amine head group. The interaction energy calculations for each residue in the binding site suggested

that Asp 105 [3] plays a much more significant role in binding antagonists with a charged tertiary amine head group than the ones with a quaternary ammonium head group. We performed quantum mechanical calculations at the Hartree-Fock level with 6-31G** basis set. The charge density at the atom centers suggested that while there is an extensive delocalization of positive charges onto the surrounding methyl groups in quaternary ammonium, the positive charges are more concentrated on the protonated hydrogen in a tertiary amine, suggesting the relative importance of a salt bridge in binding the tertiary antagonists. In the predicted structure, there is a salt bridge between the terminal oxygens of the carboxylate group of Asp105 and the QNB amine group with a distance of 2.7 Å, and between Asp 105 and pirenzepine with a distance of 2.9 Å (Fig. 11 and Fig. 12). This explains well the experimental results, where the Ala mutation of Asp 105 caused a 6000 fold reduction in QNB binding, while it only caused a 100 fold reduction in NMS binding (11). In the latter case, the aromatic residues surrounding the quaternary ammonium head group play a more significant role in NMS binding through cation-π interactions. We predict that the Asp 105 to Ala mutation will also have a significant effect on pirenzepine binding.

Due to the hydrophobic nature of the antagonist binding sites, we conducted quantum mechanical geometry optimization (with B3LYP/6-31g** basis set) to evaluate the likelihood of proton transfer from QNB to Asp 105 [3] using their docked geometry as the starting structure (Fig. 23). The results suggest that in the environment of vacuum (dielectric constant =1), cyclohexane (dielectric constant=2.0), and benzene (dielectric

constant = 2.3), the proton will transfer from the QNB tertiary amine to the carboxylate group of Asp 105 [3], whereas in methanol (dielectric constant = 33.6) and in water (dielectric constant = 80.4), the proton will remain on QNB, forming a salt bridge with Asp 105 [3]. The QNB head group is surrounded by aromatic residues, which might be mimicked by a benzene rich environment. In addition, benzene has a dielectric constant of 2.3, which is very similar to the averaged dielectric constant of amino acids (average value = 2.5). Therefore, our results suggest that proton transfer might occur upon QNB binding. Antagonists with a quaternary ammonium head group lack the proton that can be transferred to Asp 105 [3] upon binding, which also explains the relative importance of Asp 105 [3] in tertiary antagonist binding that we mentioned earlier.

The Aromatic Cage that Stabilizes the Ligand-Receptor Complex

It has been shown experimentally that the benzene ring at the end of the NMS tropic acid side chain was essential for high-affinity NMS binding (8). The predicted structure enabled us to examine, at the atomic level, how this benzene ring is stabilized by the surrounding aromatic residues including Tyr 106 [3], Trp 157 [4], and Trp 378 [6] (Fig 9C). As we described above, the Ala substitution mutations of both Tyr 106 [3] and Trp 157 [4] significantly reduced NMS binding. In a separate study (12) where Trp 378 [6] was mutated to Ala, the NMS binding was reduced 115 fold, while its Phe mutation had negligible effect (1.8 fold). This confirmed the importance of the aromatic nature of this residue. In the predicted NMS binding site, the NMS benzene ring, Tyr 106 [3], Trp 157 [4], Trp 378 [6], and additional aromatic residues including Phe 197 [5] and TM6 Tyr 381 [6], form an aromatic cage that contributes to the stability of the receptor-NMS complex.

The aromatic-aromatic interactions observed in our predicted binding site are very similar to the ones commonly observed in many protein crystal structures, where the positively charged hydrogen atoms at one aromatic ring lie closely to the negatively charged carbon atoms at another aromatic ring, forming favorable edge-face or parallel-displaced interactions (26, 27).

The predicted binding sites of QNB and 4-DAMP show that similar aromatic-aromatic interactions are also critical for QNB and 4-DAMP binding, both of which contain two benzene rings that interact with the M1 aromatic residues. The available mutagenesis results and the effect on 4-DAMP and QNB binding confirmed the importance of these aromatic residues in the binding site. The known mutations include Trp 378 [6] to Ala which reduced 4-DAMP binding 288.2 fold (12), Trp 157 [4] to Phe mutation which reduced 4-DAMP binding 131.9 fold (12), Trp 157 [4] to Ala mutation which reduced QNB binding 330 fold (10), and Tyr 106 [3] to Ala mutation which reduced QNB binding 31 fold (11). All these residues mentioned above are within the 4 Å binding sites of 4-DAMP and QNB. We predict that the mutations of these aromatic residues should affect the binding affinity of pirenzepine as well.

Role of Aromatic Residues on EC2 Loop and Extracellular End of TM Helices

From examining the binding site of 4-DAMP where its terminal two benzene rings are solely responsible for its high affinity binding at the opposite side of its quaternary ammonium group, we observed that Trp 164 near the extracellular end of TM4, Tyr 179 on the second extracellular (EC2) loop, and Phe 391 near the extracellular end of TM 6 also

participate in the aromatic network, extended through the upper benzene ring of the antagonist 4-DAMP (Fig. 24) with Trp 164 [4] and Phe 391 [6] within 5 Å to 4-DAMP. The extended aromatic network explains why the addition of an extra benzene ring at the end of the antagonist increased ligand binding affinities significantly (8).

In the QNB docked structure, the above aromatic residues do not lie immediately next to QNB. In the bovine rhodopsin crystal structure, the EC2 loop forms the roof of the binding site. We describe the EC2 conformation in bovine rhodopsin as in its closed form. We conducted additional annealed dynamics to further optimize the EC2 loop, allowing the residues of the EC2 loop, the terminal 4 residues at the extra-cellular end of TM4 and TM5, and the ligand to move, while keeping all other residues fixed. We observed that the EC2 loop moved closer to the binding site and curved similarly to the EC2 loop in the bovine rhodopsin crystal structure. Fig. 25 compares the original predicted structure (with an open EC2 loop built by the MODELLER homology modeling program (36)) and the EC2 closed structure. The closure of EC2 loop brought Tyr 179 [EC2] and Trp 164 [4] to within 5 Å of QNB, where these aromatic residues participate in the aromatic network that contributes to stabilizing the QNB binding site. Further mutagenesis to study the effect on QNB and 4-DAMP binding by the Ala mutations of these residues, especially Trp 164 [4], which is also conserved across M1-M5, will help to confirm the role of these aromatic residues in antagonist binding.

The Role of Asn 382 of TM6 in Antagonists Binding

In the predicted NMS binding site, Asn 382 [6] forms a hydrogen bond with the hydroxylmethyl group in the tropic acid side chain of NMS (Fig. 9C). The Ala substitution mutation of Asn 382 reduced NMS binding by more than 1000 fold, indicating its importance in NMS binding.

In the pirenzepine binding site (Fig. 13), Asn 382 [6] forms a hydrogen bond with the lactam group of pirenzepine. This hydrogen bond and the salt bridge with Asp 105 [3] should be important in anchoring pirenzepine from its opposite ends in M1. The Asn 382 [6] to Ala mutation had strong effects on pirenzepine binding, reducing its binding affinity by 7700 fold in M1 (41) and by 842 fold in M3 (42). Our results correspond well to these mutagenesis data.

In the QNB binding site (Fig. 12), Asn 382 [6] forms a hydrogen bond with the hydroxy group of QNB. The Ala mutation of Asn 382 [6] caused a 90 fold reduction in QNB binding. QNB has two terminal aromatic rings while NMS has only one aromatic ring that could interact with the aromatic residues in the binding site. This explains why the Ala substitution mutation of Asn 382 [6] had a stronger effect on NMS binding (reducing it by more than 1000 fold) than on QNB binding affinity (reducing it by 90 fold).

The Binding Site of the Ester Moiety of Antagonists

For all the antagonists except for pirenzepine, the carbonyl oxygen of their ester moiety is hydrogen-bonded to the hydroxy group of Ser 109 [3] at the predicted binding sites.

However, calculation of the interaction energy of the ligand with the residues in the binding site (within 5Å) suggested that this is a relatively weak hydrogen bond. For example, the calculated non-bond interaction energy is -1.3 kcal/mol between NMS and Ser 109 [3] and -2.0 kcal/mol between QNB and Ser109 [3]. This is a weak interaction compared to the interaction energy of -6.8 kcal/mol between NMS and Asn 382 [6], which explains why the Ala mutation of Ser 109 [3] did not have a significant effect on NMS and QNB binding (11). Because an ester is a relatively weak hydrogen bond acceptor (28), we suspect that anchoring the two ends of an antagonist might be sufficient for its high affinity binding. In fact, in the crystal structure complex of carbamylcholine binding to Acetylcholine-binding Protein (21), a water-soluble homolog of the ligand binding domain of nicotinic acetylcholine receptors, there is no hydrogen bond between the receptor and the carbamylcholine ester moiety (Fig. 26). This view is also supported by other results (8). Ward 1999 measured the binding affinities of the M1 antagonists diphenylacetyltropine and benztropine; both are atropine derivatives (NMS variants) where the ester carbonyl group is present in diphenylacetyltropine but removed in benztropine. Ward showed that removing this carbonyl group actually increased the binding affinity of diphenylacetyltropine by 20 fold (8).

Agonist Binding Site in the M1 Receptor

The Binding Site of the Charged Quaternary Ammonium Head Group: Comparison with the Nicotinic Acetylcholine Binding Site

Acetylcholine is the endogenous agonist for the muscarinic acetylcholine receptors and has a binding affinity of 17 μM for M1. Oxotremorine-M is an agonist with a binding affinity of 9.9 μM (12). In the predicted binding sites (as shown in Fig. 14 and Fig. 15), the quaternary ammonium head group of both acetylcholine and oxotremorine-M are stabilized by the negatively charged Asp 105 [3] as well as the surrounding aromatic residues Tyr 381 [6], Tyr 404 [7], and Tyr 408 [7] through cation-π interactions. These residues were shown experimentally to be important in acetylcholine binding. For example, acetylcholine mustard, an analogue of acetylcholine in which its headgroup was replaced by a chemical reactive aziridinium moiety, selectively labeled Asp 105 (29). In addition, the Ala substitution mutations of Tyr 381 [6], Tyr 404 [7], and Tyr 408 [7] each reduced acetylcholine binding affinity by about 30 fold (8, 10). Note that the distances between the acetylcholine head group and the surrounding aromatic rings (Fig. 14) are very similar to the ones observed in the crystal structure (21) of the nicotine acetylcholine receptor binding site (Fig. 26). Similar distances between the quaternary ammonium group and the phenyl rings were also observed in the binding sites of antagonists (NMS and 4-DAMP) that have a quaternary ammonium head group (Fig. 9C and Fig. 11).

Binding site of the Ester Moiety in Agonists Leads to Breaking the Interhelical Hydrogen Bond between TM 3 and TM 6

In the predicted binding site in M1, the ester moieties of acetylcholine and oxotremorine-M are hydrogen bonded to both Tyr 381 [6] and Ser 109 [3] (Fig. 14 and Fig. 15). The Ala

mutation of Tyr 381 [6] and Ser 109 [3] reduced acetylcholine binding by 30 and 10 fold, respectively (8, 11). In addition, it has been shown (8) that the hydroxy group of Tyr 381 [6] is important in acetylcholine binding in M1 because the Phe mutation of this residue significantly reduced acetylcholine binding in a magnitude similar to an Ala mutation (while having negligible effects on NMS binding). In M2, the equivalent mutation (Tyr403 to Phe) also had a differential effect on agonist and antagonist binding. This mutation reduced the NMS and QNB affinities to 2.5 and 1.7 fold, respectively, but reduced acetylcholine affinity 280 fold (47). To test this differential effect on agonists and antagonists binding in the predicted M1 structure, we mutated Tyr 381 to Phe at the NMS, QNB, and acetylcholine binding sites and recalculated their binding energy. We observed a small decrease (6%) in binding energy for NMS and no change in binding energy for QNB. However, there is a significant decrease (44%) in binding energy for acetylcholine, which parallels the experimental data.

In the predicted structure we found that there is an interhelical hydrogen bond (2.9 Å) between Tyr 381 [6] and Ser 109 [3]. When the agonists (acetylcholine and oxotremorine-M) bind to this site, they break this interhelical hydrogen bond by wedging in between and compensate by making a hydrogen bond between the ester moiety and the Tyr381 [6] (Fig. 27). This interhelical hydrogen bond is not broken when an antagonist binds to this site (Fig 27). This breaking of interhelical hydrogen bond on agonist binding might contribute to receptor activation by weakening the intra-molecular contacts between TM 3 and TM 6.

The Internal Torsional Conformation of the Docked Aceytlcholine :

The docked conformation of acetylcholine at the predicted binding site is a *gauche* conformation with an NCCO torsion angle of 59°, and it is the lowest energy conformation after the extensive internal conformational sampling of the Hierdock simulated annealing procedures used for predicting the binding site. The predicted torsion angle corresponds extremely well to the experimentally observed value (an NCCO torsion angle of 60°) in a transferred NOE NMR study of the acetylcholine analogue S-(+)-acetyl-β-methyl choline bound to the ground state of the M2 receptor (30). The same experiment also suggested that the conformation of bound acetylcholine in the activated state of the receptor is *trans* with an NCCO torsion angle of 132°. This result, together with the rotational orientation of TM5 and the inter-helical contacts involving Leu 116 [3], Phe 374 [6], and Asn 414 [7] that selectively stabilize the receptor ground state (Fig. 2D), suggest that our predicted structure might correspond to the inactive state of the receptor.

The Role of TM 5 in Agonist and Antagonist Binding

In the rotational optimization of the helical orientations (discussed in the Methods section and shown in Fig. 39) there are two possible orientations found for TM5. They are 0 degrees and 90 degrees. In the current structure discussed, we have taken the 0 degrees rotation structure for TM5 because it is the orientation that has the lowest energy in rotational energy scans for TM 5. In this structure, TM5 is at the farthest end of the binding site with Ala 193 [5] pointing towards the binding site. This rotational orientation of TM5 corresponds well to the experimental results reported by Allman 2000, which, out

of the mutations of Ile 188 [5] through Ala 196 [5], only that of Ala 193 [5] to Cys reduced the binding affinity by 3.6 fold (9). In addition, the sulfhydryl reagent N-trimethyl-2-aminoethyl methanethiosulfonate reacted selectively with Ala 193 [5] when it was mutated to Cys and the residues on both sides of Ala 193: Phe 190 [5] → Cys and Thr 192 [5] → Cys (9). Allman et al 2000 also proposed that an agonist induced rotation of TM5 might accompany the receptor activation (9). Interestingly, we observed a second energy minimum in TM5 360° energy rotational scans at the orientation where TM5 is rotated 90° clockwise looking from the extracellular side. In this orientation, Thr 192 [5] points toward the binding site, which is exactly the TM5 orientation of the active state suggested by Allman (9).

Role of Trp 157 [4] in Agonist Binding

Trp 157 [4], the residue whose Ala substitution significantly affected antagonist binding, also reduced acetylcholine binding 90 fold. In the predicted structure, Trp 157 [4] is within a 5.5 Å distance to the bound acetylcholine (Fig. 14) and is a part of the aromatic network. The mutation of Trp 157 [4] might affect the packing of the binding site residues including its immediate neighbor Tyr 106 [3] that interacts with acetylcholine. As we discussed earlier, different patches of residues of TM5 affected agonist or antagonist binding. At the TM5 rotational orientation of 90 degrees presumably corresponding to the orientation at the activated state, Thr 189 [5], Thr 192 [5] and Ala 196 [5] that selectively affected

acetylcholine binding are located immediately next to Trp 157 [4], suggesting a possible role of Trp 157 [4] in acetylcholine binding in the active state of the receptor.

Comparison of Agonist and Antagonist Binding Sites

By comparing the structures of M1 agonists and antagonists, we notice that acetylcholine, oxotremorine-M, and many other agonists can be superimposed to the antagonist NMS and 4-DAMP because they share the quaternary ammonium head group and the ester moiety. However, acetylcholine lacks the two benzene rings in 4-DAMP and the hydroxylmethyl and benzene ring at the end of the NMS tropic acid side chain that can form aromatic-aromatic interactions or hydrogen bonds with the binding site residues of TM 3, 4, 5, and 6. This is probably how antagonists keep the receptors from activation by locking these TM helices in place, while agonists allow their conformational flexibility required for activation. This view was suggested earlier (31). The antagonist nature of benzilylcholine (8) also supported this view. In benzilylcholine, two benzene rings and a hydroxy group (as seen in QNB) replaces the terminal methyl group (next to the ester) of acetylcholine, which switches it from an agonist to an antagonist. Fig. 28 summarizes the binding regions of antagonists and agonists used in this study. The binding regions for agonists and antagonists overlap between TM 3, 6, 7 while the antagonists binding region extends to TM 4 and TM 5.

Comparison of the Calculated Ligand Binding Energies and Experimental Dissociation Constants

We calculated the binding energies of all the 10 ligands used in this study and compared the results with their experimental dissociation constants –log (Kd), -log (Ki), or –log (IC50). The molecules selected represent a diverse set of ligands, including both agonists and antagonists. Fig. 28 show that the calculated relative binding energies correlate well with experimentally measured binding affinities. The R square values for antagonists, agonists, and all ligands are 0.84. It should be noted that the calculated binding energies do not include the room temperature effects, or explicit entropic terms.

Conclusion

The predicted 3D structure of the human M1 receptor, and the predicted binding sites of 10 agonists and antagonists, both based on first principles, provide a sound structural basis for the measured ligand binding affinities and mutagenesis data. The calculated relative binding energies also correlate well with measured binding affinities. In addition, the predicted binding sites provide a structural basis to better explain the large reduction in ligand binding affinity and signaling efficacy by mutations of Trp 157 and Pro 159 of TM 4 for which prior models differed. The predicted structure explains the differential effect of Tyr 381 [6] to Phe mutation on agonist and antagonist binding, as well as the stronger reduction in binding affinity by Asp 105 [3] to Ala mutation for antagonists with a tertiary amine head group. It also explains why the Asn 382 [6] to Ala

mutation affected NMS more than QNB. The predicted structure shows the importance of binding site aromatic residues and illustrates, in atomic detail, the cation-π interactions and aromatic-aromatic interactions critical to ligand binding in M1. Several new mutation candidates were suggested, including Trp 164 at the extracellular end of TM 4 (for its effect on the binding affinities for antagonists with two terminal benzene rings) and additional aromatic residues (Phe 197 [5], Tyr 179 [EC2], and Phe 391 [6]) that are part of the aromatic network stabilizing the antagonist binding sites. The detailed structure of M1 and its binding sites provides a basis for structure-based drug design. This study and several successful recent studies in other GPCRs justify carrying out similar studies to predict the structures and ligand binding of the other four muscarinic receptor subtypes, which should, in turn, improve our understanding of the origins of subtype specificities in muscarinic receptors.

Chapter 3

COMPUTATIONAL METHODS

Structure Prediction of GPCR: The MembStruk Method

The Membstruk method (version 4.0) was used to predict the M1 structure with each step listed in the following section. A detailed description of the Membstruk method (version 3.5) used here has been published (14). In Membstruk 4.0, we have made advances in generating a large number of well packed conformations with additional criteria including total all-atom forcefield energy, number of interhelical hydrogen bonds and salt bridges in determining the rotational orientation of the transmembrane (TM) regions. All energy and force evaluations used the DREIDING force field (32) and CHARMM22 charges (33) for the protein. In this chapter we only detail the advances in the MembStruk 4.0 method from its previously published version MembStruk 3.5 (14). The various steps as applied to prediction of the M1 structure are:

TM Prediction (TM2ndS).

We predicted each TM region by using alignments of 51 sequences that have sequence similarities to M1 varying from 99% to 24% using the BLAST search. The alignment includes 45 M1-M5 sequences from various species and 4 sequences from histamine

receptors and 2 from octopamine receptors. M1-M5 receptors have more than 40% overall sequence identity and 60% sequence identity in the TM regions. Histamine and octopamine receptors were identified by BLAST to have the highest homology to M1 outside of the muscarinic receptor family. Histamine and M1 receptors have 29% overall sequence identity and 37% sequence identity in the TM regions, while octopamine and M1 receptor have 24% overall sequence identity and 35% sequence identity in the TM regions. By including sequences with lower homology, we improved the identification of the boundary between TM 2 and 3 (the EC1 loop), which is highly hydrophobic in all M1-M5. Then, we calculated the average hydrophobicity for every residue position over all the sequences in the multiple sequence alignment and averaged over window sizes of 12 to 20 residues. The baseline for this profile served as the threshold value for determining the TM regions (Fig 1). Because it is possible that the actual length of the helix could extend past the membrane surface, we carried out an additional capping step as described in Trabanino 2004 (14), based on properties of known helix breaker residues, to further refine where each predicted TM region ends.

Optimization of the Translational Orientation of the Helices.

TM2ndS also predicted the position of maximum hydrophobicity called the "hydrophobic center" for each helix. The seven canonical α-helices were built with the predicted TM sequences and the helical axes were positioned based on the 7.5-Å density map of frog

rhodopsin (34). Relative translational orientation of the helices was optimized by fitting the hydrophobic centers of the TM helices to a common plane.

Optimization of Helical Bends and Kinks

We constructed canonical helices for the predicted TM segments and optimized the individual helices using energy minimization followed by Cartesian molecular dynamics at 300K for 500 ps. This procedure optimized the bends and kinks in each helix.

Optimization of Rotational Orientation of the Helices

Step 1 Hydrophobic moment: The rotational orientation of the helices was determined according to the hydrophobic moment of the middle portion of each helix (about its hydrophobic center) to point the moment of each helix towards the lipids.

Step 2 Monte Carlo optimization (*RotMin*): In this procedure, each of the seven TMs was optimized through a range of rotations one at a time while the other six helices were re-optimized in response. After each rotation, the side-chain positions of all residues for all seven helices in the TM region were optimized using SCWRL (35). The potential energy of the rotated helix was then minimized in the field of all other helices. This procedure was carried out for a grid of rotation angles (every 5° for a range of ±25°) to determine the optimum rotation for each helix. Once we rotated the helix to its optimum rotation, we kept it fixed while allowing each of the other six helices to be rotated and optimized. The

optimization of these six helices was done iteratively until the entire grid of rotation angles was searched. This method optimized the rotations locally.

Step 3 Energy Scan: This step is the recent refinement from MembStruk 3.5 and is included only in Membstruk 4.0. We took the structure generated from step 2 as the initial structure and rotated each individual helix while fixing the other six helices. We rotated the helix in 5° intervals for 360° and calculated the energy of the rotated helix after reassigning the side chain conformation (with SCRWL) and minimizing the energy of the entire protein (35). We also calculated the number of interhelical hydrogen bonds and salt bridges to identify favorable inter-helical interactions. Step 3 allowed us to identify additional low-energy rotation orientations for each helix assuming Step 1 and 2 predicted good rotational orientations for other helices. We conducted energy scans iteratively until we reached a structure where all the seven helices are within 15° of its lowest energy minimum in 360° scans. The scanning results are shown in Fig. 30-43 where the 0° rotational orientation represents the final conformation of each helix. Because Step 2 and 3 are more important for optimizing helix 3, which is less amphipathic due to its center location, we docked ligands to both the 0° conformation and the conformation of its second lowest energy minimum where TM3 was rotated 20° clockwise from its lowest minimum (at 0°). These two structures are very similar. While most ligands dock very well to the 0° structure, the antagonist NMS and its derivatives, having bigger head groups than acetylcholine and other ligands, do not fit as well in the 0° structure as they fit in the 20° structure. Therefore, the 20° structure was selected and reported here.

Loop Building and Optimization of the Final Model

The loops were added to the helices using the loop builder in the MODELER homology modeling program (36). We omitted modeling the amino terminal before TM1 and carboxyl terminal region after TM7. The third intracellular loop between TM 5 and TM 6 was truncated to 25 residues with the residues 225 to 353 deleted (the dLoop mutation) because such large loops are quite flexible, leading to a multiplicity of conformations with similar energies. It has been shown experimentally that the d-Loop mutation had little effect on ligand-binding affinities (37) and should have little effect on the structure of the TM region. Disulfide linkages were created between Cys 98 [TM 3] – Cys 178 [EC2 loop]. We then performed side chain replacement with SCWRL and optimized the final structure by using conjugate gradient minimization of all atoms in the structure.

Functional Prediction of GPCRs: The HierDock Protocol

The ligands used in this study were described with the DREIDING FF and Gasteiger charges (38). The HierDock 2.0 procedure (13) was used. The ligand-screening protocol followed a hierarchical strategy for examining ligand-binding conformations and calculating their binding energies. This method has been tested for various GPCRs (13-20), outer membrane protein A (39), and globular proteins (40-43). The HierDock protocol is

described in detail in these references. In brief, the parameters used for ligand binding site predictions in M1 are as follows:

Scanning the Entire Receptor for Binding Sites

The void space in the entire receptor was mapped using the SPHGEN DOCK 4.0 utility (29). The entire set of spheres was divided into 27 regions for scanning and locating the binding site of ligands without any prior bias. Five representative ligands including NMS, QNB, 4-DAMP, acetylcholine, and tetramethylammonium were used for scanning. In each of these regions we carried out DOCK 4.0 to generate and score 500 conformations, of which 50 (10%) were selected using a buried surface area cutoff of 85% and energy scoring from DOCK 4.0. In Step 2, the 50 best conformations selected were subjected to all-atom minimization keeping the protein fixed but the ligand movable. The 5 best scoring conformations based on the potential energy of the ligand in the protein were selected. We then examined which boxes contained these top 5 hits and selected the best boxes based on results from all the ligands. We merged the spheres in the top two scoring boxes as the putative binding region for the next step (Fig. 8).

Optimizing the Binding Sites

Once the plausible binding regions were determined, we docked all the ligands shown in Fig. 7 into the putative binding region using the HierDock 2.0 protocol. The best 5 bound

structures for each ligand/M1 receptor complexes were chosen. We then optimized this structure using the SCREAM side chain replacement program (Kam, Vaidehi, and Goddard, unpublished) to reassign side chains for the residues within 4 Å of the ligand. Because the receptor structure is kept fixed in HierDock, we performed simulated annealing dynamics to optimize the binding sites, allowing both the receptor side chains and the ligands to move. Specifically, we ran annealed dynamics for a total of 120 ps cycling between 300K and 700K for 300 cycles for all docked ligands, allowing the ligands and the side chain atoms within 5Å from the ligands to move freely while keeping the main chain atoms fixed. The ligand-receptor complexes of N-Methylhomatropine and N-Methylacetyltropine were derived by matching the ligands to the optimized N-Methylatropine / M1 complex conformation.

Binding Energy Calculation

We calculated the binding energy (BE) of each ligand using the equation BE = PE (ligand in protein) – PE (ligand in solvent) as the difference between the potential energy (PE) of the ligand in the protein and the potential energy of the ligand in water. The energy of the ligand in water was calculated by using DREIDING FF and the surface Generalized Born Continuum solvation method (45).

References

1. Eglen, R. M., Choppin, A., & Watson, N. (2001) *Trends Pharmacol. Sci.* **22,** 409-414.

2. Bymaster, F. P., McKinzie, D. L., Felder, C. C., & Wess J., (2003) *Neurochem. Research* **28**, 437-442.

3. Messer, W. S., Jr., (2002) *Curr. Top. Med. Chem.* **2,** 353-358.

4. Greenlee, W., Clader, J., Asberom, T., McCombie, S., Ford, J., Guzik

, H., Kozlowski, J., Li, S., Liu, C., Lowe, D., & Vice, S. *et al.* (2001) *Il Farmaco* **56**, 247-250.

5. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., *et al.* (2000) *Science* **289,** 739-745.

6. Bissantz, C., Bernard, P., Hibert, M., & Rognan, D. (2003) *Proteins* **50**, 5-25.

7. Hulme, E. C., Lu, Z. L., & Bee, M. S. (2003) *Receptors Channels* **9,** 215-228.

8. Ward, S. D. C., Curtis, C. A. M., & Hulme, E.C. (1999) *Mol. Pharmacol.* **56**, 1031-1041.

9. Allman K, Page KM, Curtis CA, Hulme EC. (2000) *Mol. Pharmacol.* **58**, 175-184.

10. Lu ZL, Saldanha JW, Hulme EC. (2001) *J. Biol. Chem.* **36**, 34098-104.

11. Lu ZL, Hulme EC. (1999) *J. Biol. Chem.* **274**, 7309-15.

12. Bourdon, H., Trumpp-Kallmeyer, S., Schreuder, H., Hoflack, J., Hibert, M. & Wermuth, CG. (1997) *J. Comput. Aided Mol. Des.* **11,** 317-332.

13. Vaidehi, N., Floriano, W. B., Trabanino, R. J., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., & Goddard, W. A., III (2002) *Proc. Natl. Acad. Sci. USA* **99**, 12622-12627.

14. Trabanino, R. J., Hall, S. E., Vaidehi, N., Floriano. W. B., Kam, V., & Goddard, W. A., III (2004) *Biophys. J.* **86**, 1904-1921

15. Freddolino, P. L., Kalani, M. Y. S., Vaidehi, N., Floriano, W. B., Hall, S. E., Trabanino, R. J., Kam, V. W. T., & Goddard, W. A., III (2004) *Proc. Natl. Acad. Sci. USA* **101**, 2736-2741

16. Kalani, M. Y. S., Vaidehi, N., Hall, S. E., Trabanino, R. J., Freddolino, P. L., Kalani, M. A., Floriano, W. B., Kam V. W. T., & Goddard, W. A., III (2004) *Proc. Natl. Acad. Sci. USA* **101**, 3815-3820.

17. Floriano, W. B., Vaidehi, N., Goddard, W. A., III, Singer, M. S., Shepherd, G. M. (2000) *Proc. Natl. Acad. Sci. USA* **20**:10712-10716.

18. Floriano, W. B., Vaidehi, N., Zamanakos, G. & Goddard, W. A., III (2004) *J. Med. Chem.* **47**, 56-71.

19. Hall, S. E., Floriano, W. B., Vaidehi, N., Goddard, W. A., III (2004) *Chem. Senses* **29**:595-616.

20. Hummel, P., Vaidehi, N., Floriano, W. B., Hall, S. E., Goddard, W. A., III (2005) *Protein Sci* **14**:703-710.

21. Celie, P. H. N., Rossum-Fikkert, S. E. V., Dijk, W. J. V., Brejc, K., Smit, A. B., & Sixma, T. K. (2004) *Neuron* **41**, 907-914.

22. Lu, Z. L., Curtis, C. A., Jones, P.G., Pavia, J., & Hulme, E. C. (1997) *Mol. Pharmacol.* **51**:234-241.

23. Hulme, E. C., & Lu, Z. L. (2000) *J. Biol. Chem*. **275**, 5682-5686.

24. Huang, X. P., Nagy, P. I., Williams, F. E., Peseckis, S. M., & Messer, W. S., Jr., (1999) *Br. J. Pharmacol.***126**:735-745.

25. Hulme, E. C., Lu, Z. L., Saldanha, J. W., & Bee, M. S. (2003) *Biochem. Soc. Trans.* **31**, 29-34.

26. Bruley, S. K., & Petsko, G. A. (1985) *Science* **229**, 23-28.

27. Guvench, O., & Brooks, C. L., III (2005) *J. Am. Chem. Soc.* **127**, 4668-4674.

28. Deechongkit, S., Dawson, P. E., & Kelly, J. W. (2004) *J. Am. Chem. Soc.* **126**, 16762-16771.

29. Spalding, T. A., Birdsall, N. J. M., Curtis, C. A. M., & Hulme, E. C. (1994) *J. Biol. Chem*. **269**, 4092-4097.

30. Furukawa, H., Hamada, T., Hayashi, M. K., Haga, T., Muto, Y., Hiroa, h., Yokoyama, S., Nagasawa, K., & Ishiguro, M. (2002) *Mol. Pharmacol.* **62**, 778-787.

31. Tanczos, A. C., Palmer, R. A., Potter, B. S., Saldanha, J. W., & Howlin, B. J. (2004) *Comput. Biol. Chem.* **28**, 376-385.

32. Mayo, S. L., Olafson, B. D., & Goddard, W. A., III (1990) *J. Phys. Chem*. **94**, 8897-8909.

33. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, & D. Joseph-McCarthy (1998) *J. Phys. Chem.* **102**, 3586–3616.

34. Schertler, G. F. X., (1998) *Eye* **12**, 504–510

35. Bower, M., Cohen, F. E., & Dunbrack, R. L. Jr., (1997) *J. Mol. Biol.* **267**, 1268–1282

36. Marti-Renom, M. A., Stuart, A., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291-325.

37. Lu, Z. L., Curtis, C. A., Jones. P. G., Pavia, J., & Hulme, E. C. (1997) *Mol. Pharmacol.* **51**:234-241.

38. Gasteiger, J., & Marsili, M. (1980) *Tetrahedron* **36**, 3219-3228.

39. Datta, D., Vaidehi, N., Floriano, W. B., Kim, K. S., Prasadarao, N. V. & Goddard, W. A., III (2003) *Proteins Struct. Funct. Genet.* **50**, 213-221.

40. Datta, D., Vaidehi, N., Xu, X. & Goddard, W. A., III (2002) *Proc. Natl. Acad. Sci. USA* **99**, 2636-2641.

41. Zhang, D., Vaidehi, N., Goddard, W. A., III, Danzer, J. F. & Debe, D. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 6579-6584.

42. Wang, P., Vaidehi, N., Tirrell, D. A. & Goddard, W. A., III (2002) *J. Am. Chem. Soc.* **124**, 14442-14449.

43. Kekenes-Huskey, P. M., Vaidehi, N., Floriano, W. B. & Goddard, W. A., III (2003) *J. Phys. Chem B* **107**, 11549-11557.

44. Ewing, T. A. & Kuntz, I. D. (1997) *J. Comput. Chem.* **18**, 1175 1189.

45. Bashford, D., & Case, D. A. (2000) *Annu. Rev. Phys. Chem.* **51**,129-152.

46. Mecozzi, S., West. A. P., Jr., & Dougherty, D. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 10566-10571.

47. Vogel, W. K., Sheehan, D. M., & Schimerlik, M. I. (1997) *Mol. Pharmacol.* **52**, 1087-1094.

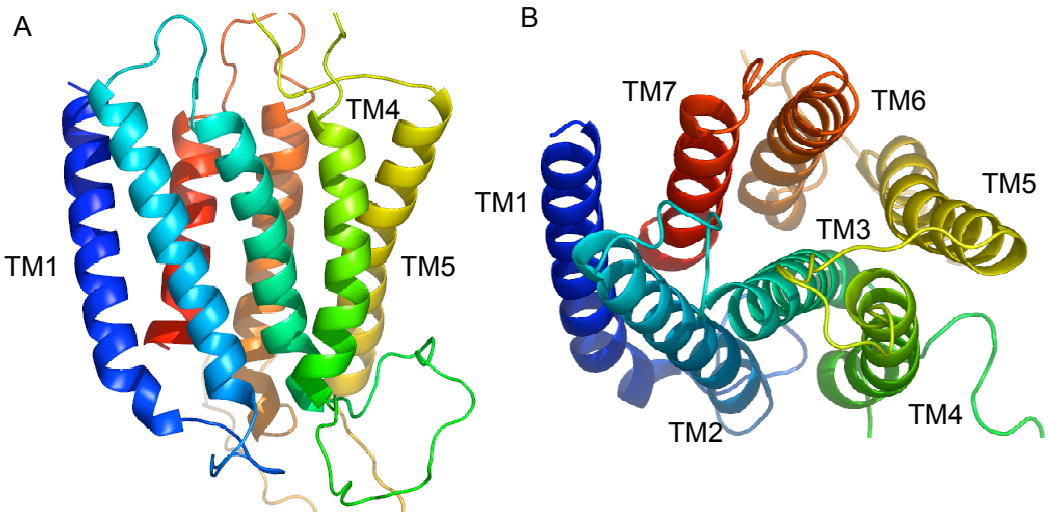


Fig. 1 (A) The side view of the predicted M1 structure (B) The top view of the predicted M1 structure.

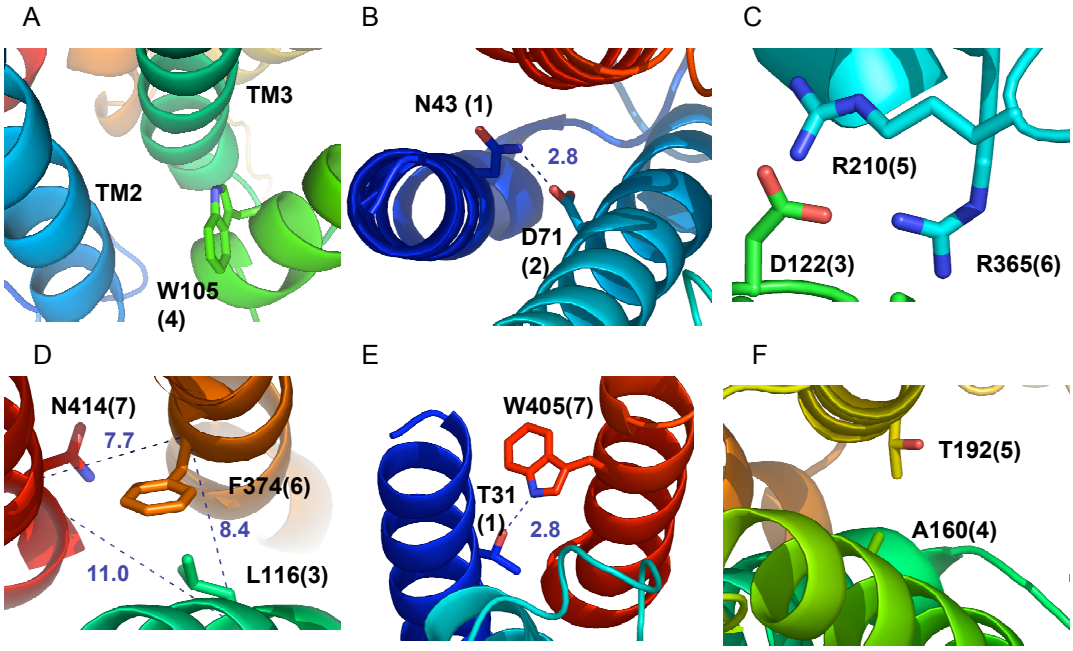


Fig. 2 Illustration of the details of the predicted M1 structure. The numbers shown in parentheses are the TM helix to which the residue belongs.

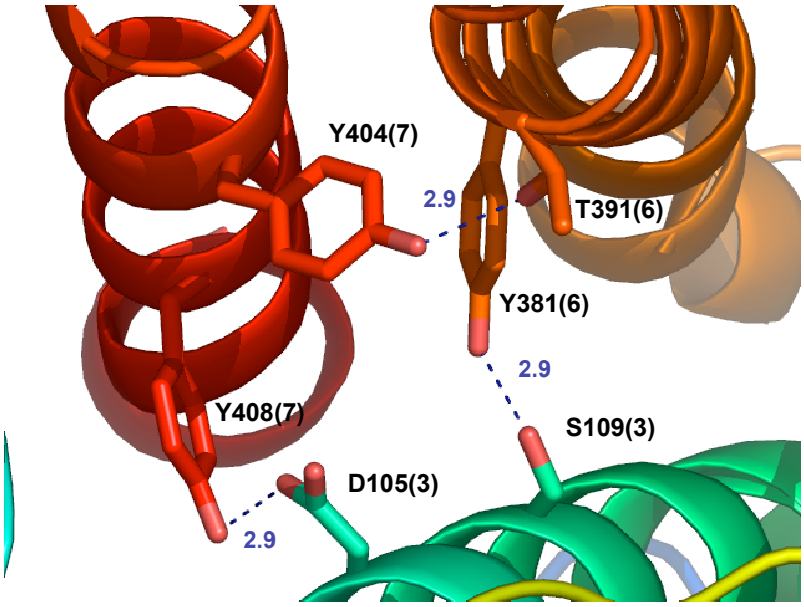


Fig. 3 The inter-helical hydrogen bonds observed at the M1 binding sites. The numbers shown in parentheses are the TM helix to which the residue belongs.

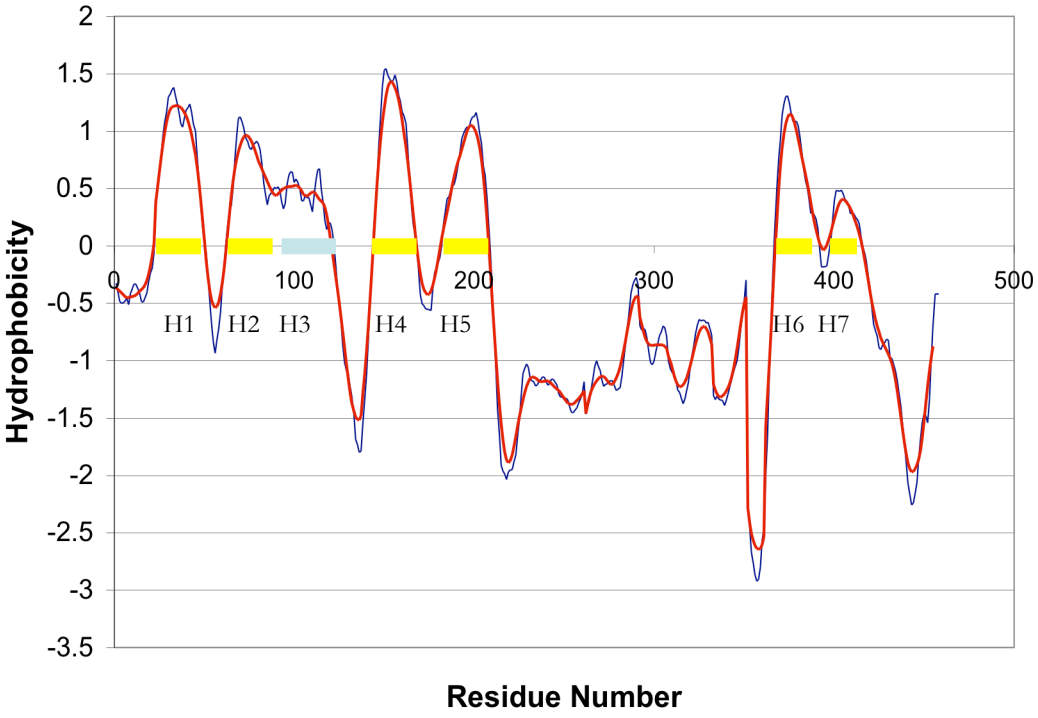


Fig. 4 The hydrophobicity profile from TM2ndS for M1 averaged over window sizes of 12 to 20 residues. 51 sequences are used in the alignment as described in the methods.

N Terminal (22)	MNTSAPPAVSPNITVLAPGKGP
TM 1 (29)	WQVAFIGITTGLLSL**A**TVTGNLLVLISFK
IC 1 (6)	VNTELK
TM 2 (31)	TVNNYFLLSLACADL**I**IGTFSMNLYTTYLLM
EC 1 (6)	GHWALG
TM 3 (28)	TLACDLWLALDYVAS**N**ASVMNLLLISFD
IC 2 (21)	RYFSVTRPLSYRAKRTPRRAA
TM 4 (23)	LMIGLAWLVSF**V**LWAPAILFWQY
EC 2 (16)	LVGERTVLAGQCYIQF
TM 5 (28)	LSQPIITFGTAMAAF**Y**LPVTVMCTLYWR
IC 3 (25)	IYRETENRARELAATFSLVKEKKAA
TM 6 (28)	RTLSAILLAFILT**W**TPYNIMVLVSTFCK
EC 3 (5)	DCVPE
TM 7 (21)	TLWELGYWLCYV**N**STINPMCY
C Terminal (43)	ALCNKAFRDTFRLLLLCWDKRRWRKIPKRPGSVHRTPSRQC

Fig. 5 The predicted TM regions and the hydrophobic centers (underlined) of each helix. The length of each region is also listed. The third intracellular loop between TM 5 and TM 6 was truncated to 25 residues with the residues 225 to 353 deleted. It corresponds to the dLoop mutation structure that was shown experimentally to have little effect on ligand binding (37).

TM Helix	Helix Bend (degrees)	Helix Tilt (degrees)
1	6.4	20.2
2	3.9	17.6
3	6.9	23.2
4	12.0	5.1
5	14.1	26.5
6	18.2	12.9
7	20.3	21.5

Fig. 6 Important structure parameters of the Membstruk structure of the human M1 receptor. The helical tilt is the angle (in degrees) between the mean helical axis and the mean plane that passes through the hydrophobic masimum of each helix. The helical bend is defined as the angle between the middle of the TM region and the top of the helix as well as between the middle and the bottom of the helix.

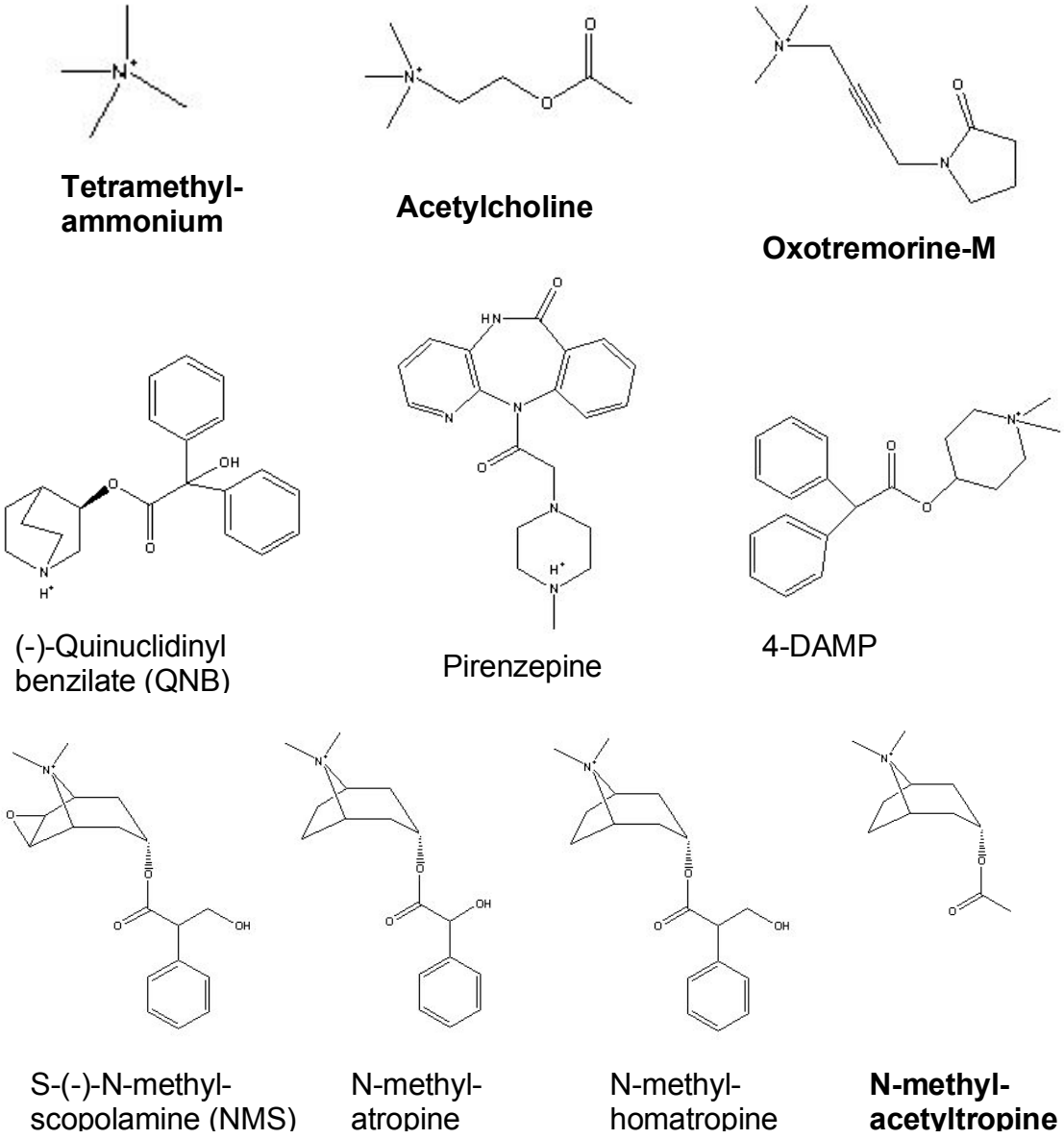


Fig. 7 The 10 ligands used in this study which include (1) agonists (in bold): tetramethylammonium, acetylcholine, and oxotremorine-M (2) antagonists with a tertiary amine head group: QNB and pirenzepine (3) and antagonists with a quaternary ammonium head group: 4-DAMP, NMS, N-Methylatropine, N-Methylhomatropine (4) unknown: N-Methylacetyltropine (in bold), which binds similarly to agonists. The biological active S-(-)-enantiomer of NMS and other atropine derivatives are used.

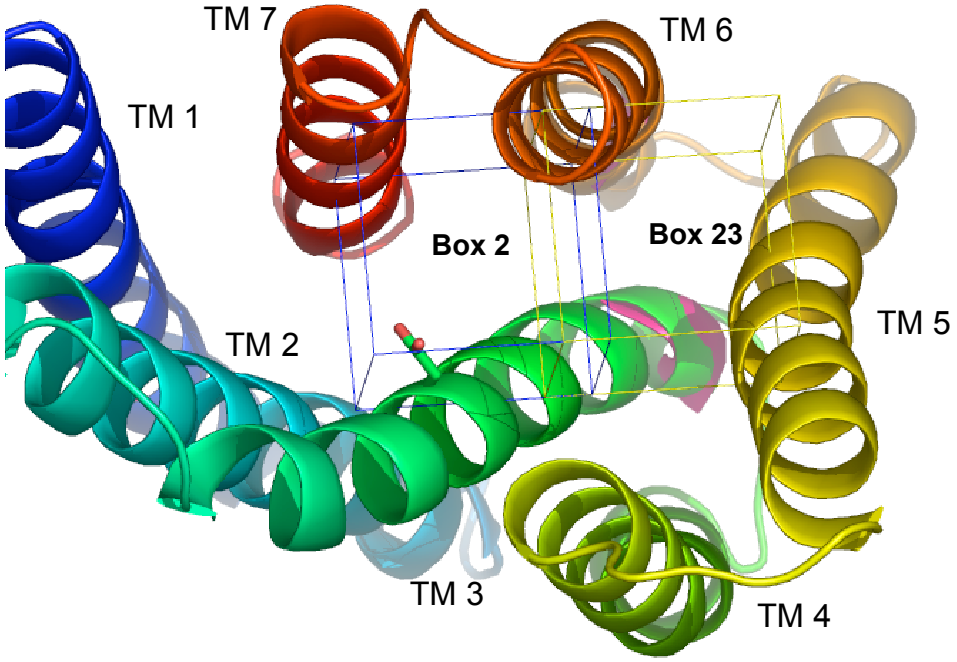


Fig. 8 The top two boxes where the ligands docked the best to. Asp 105 (3), which is located in box 2, is also shown.

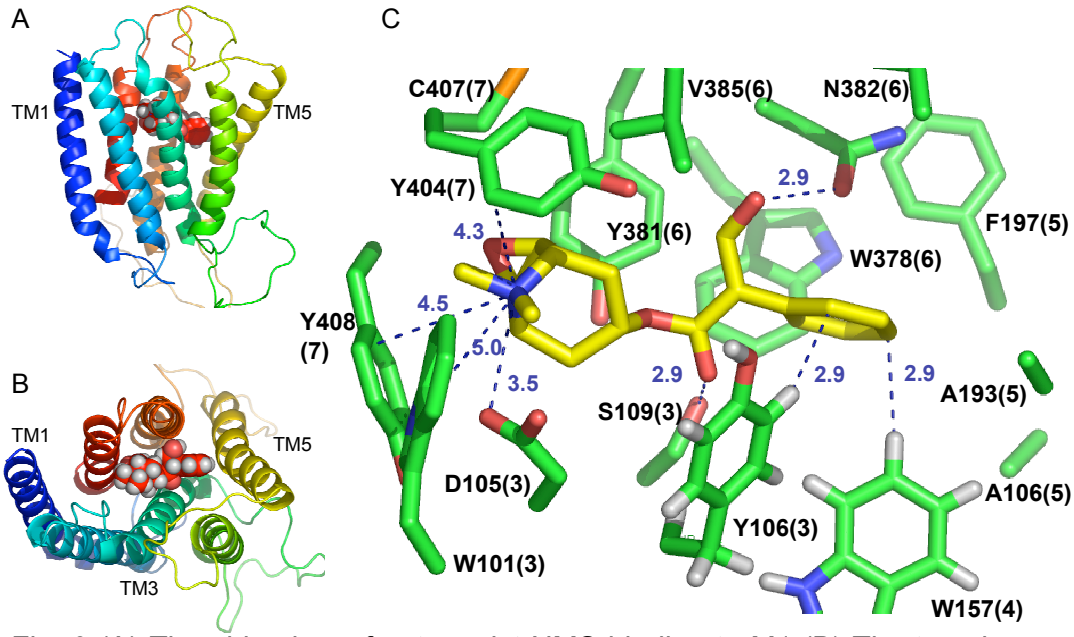


Fig. 9 (A) The side view of antagonist NMS binding to M1 (B) The top view of NMS binding to M1 (C) The 4 Å binding site of NMS. The numbers shown in parentheses are the TM helix to which the residue belongs. Distances of the important interactions between NMS and M1 binding site residues are highlighted.

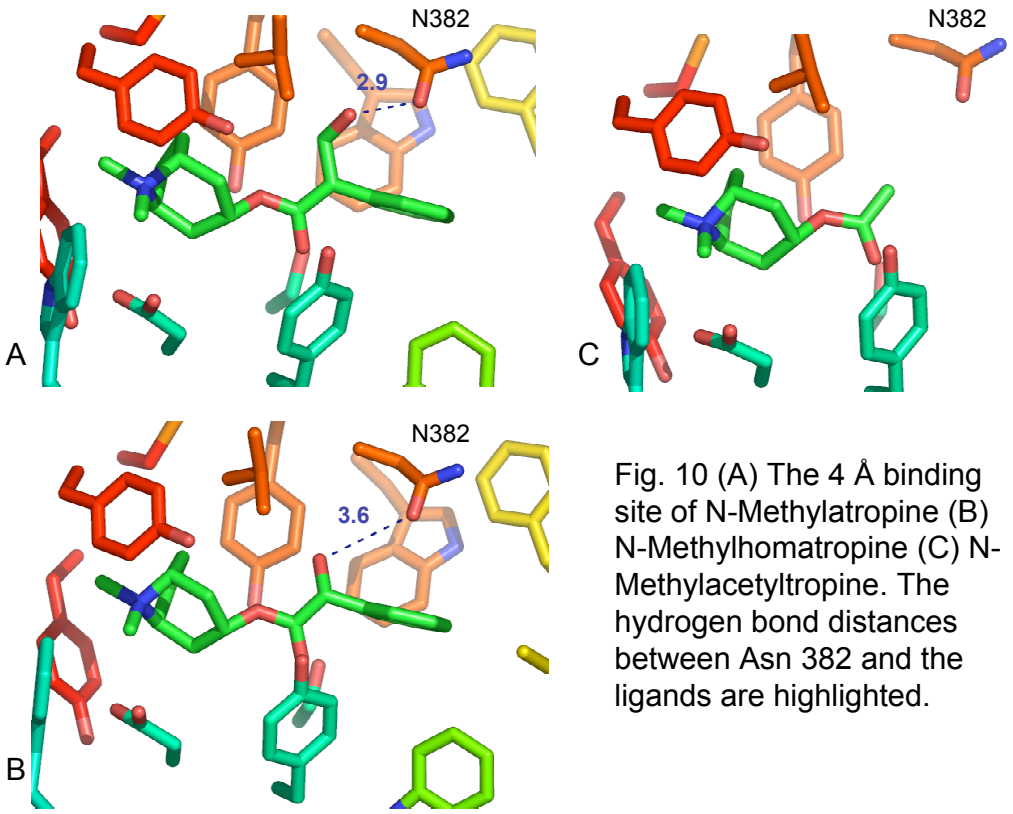


Fig. 10 (A) The 4 Å binding site of N-Methylatropine (B) N-Methylhomatropine (C) N-Methylacetyltropine. The hydrogen bond distances between Asn 382 and the ligands are highlighted.

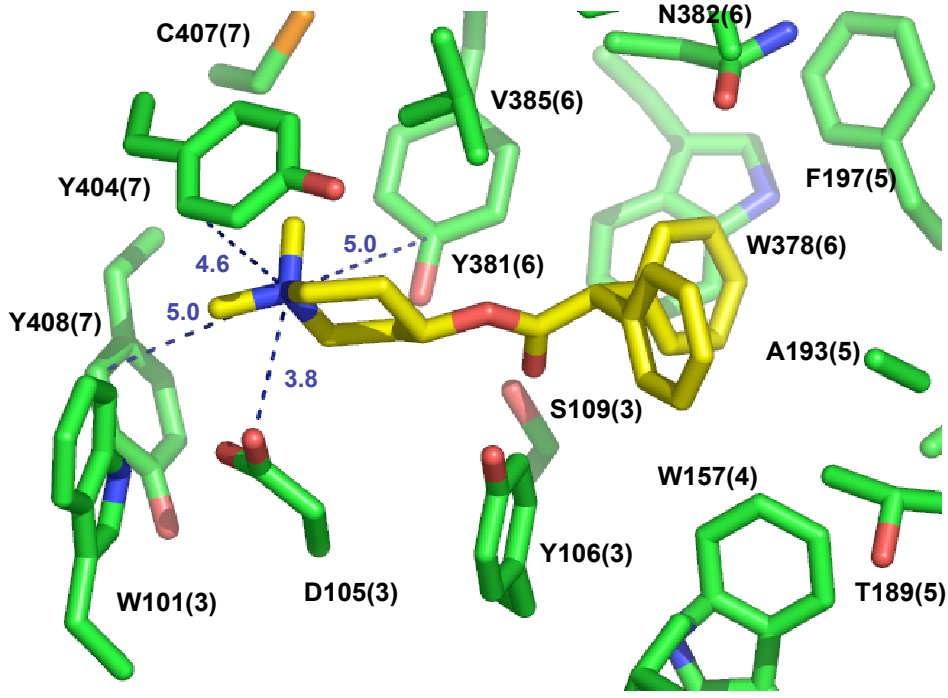


Fig. 11 The 4 Å binding site of antagonist 4-DAMP. The numbers shown in parentheses are the TM helix to which the residue belongs.

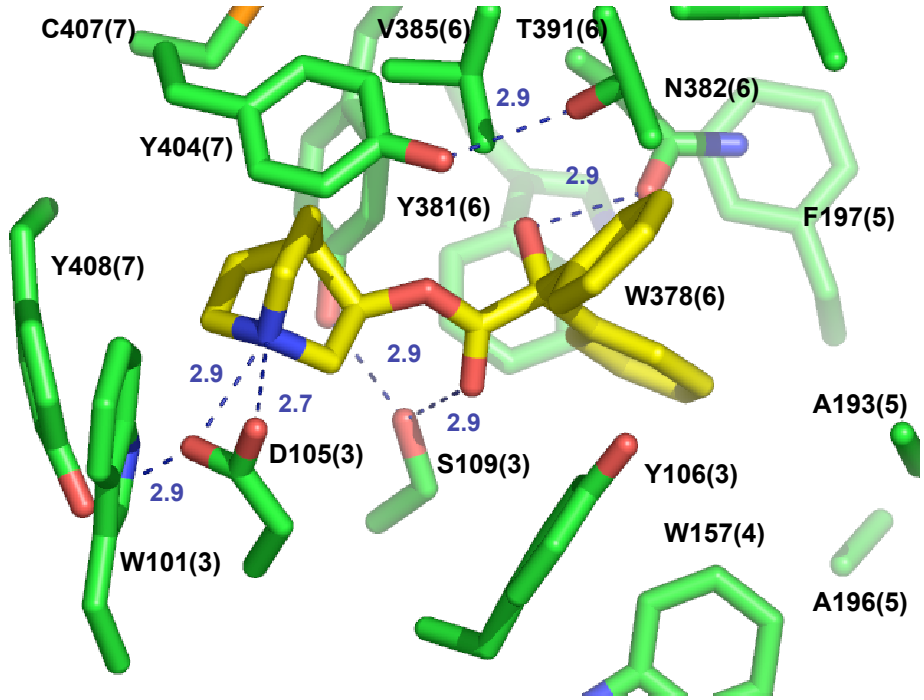


Fig. 12 The 4 Å binding site of antagonist QNB. The numbers shown in parentheses are the TM helix to which the residue belongs.

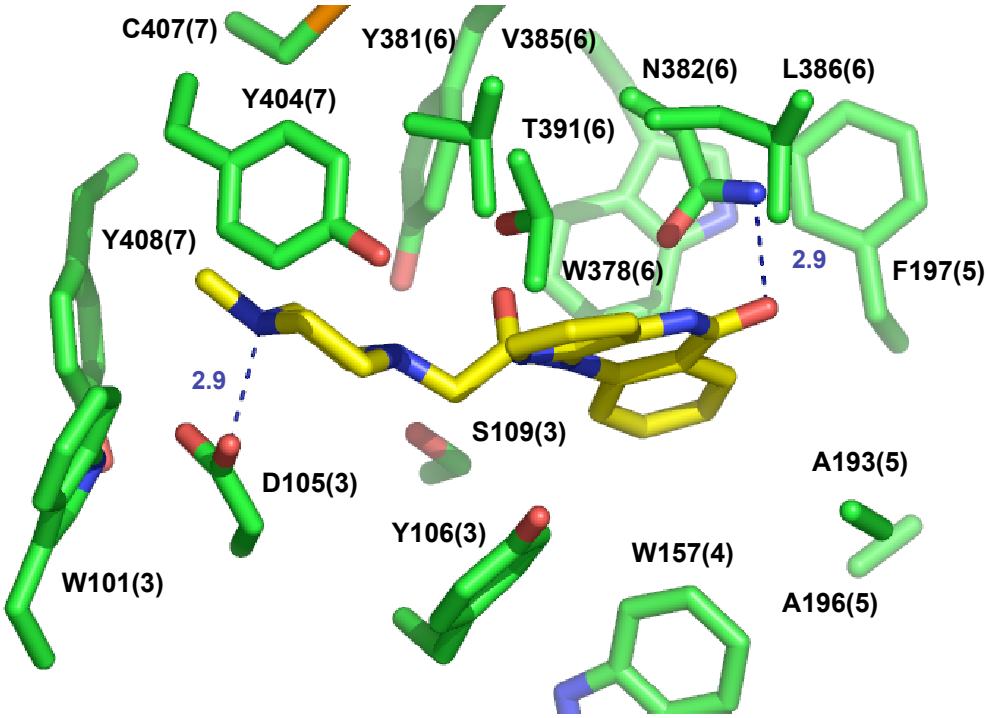


Fig. 13 The 4 Å binding site of antagonist pirenzepine. The numbers shown in parentheses are the TM helix to which the residue belongs.

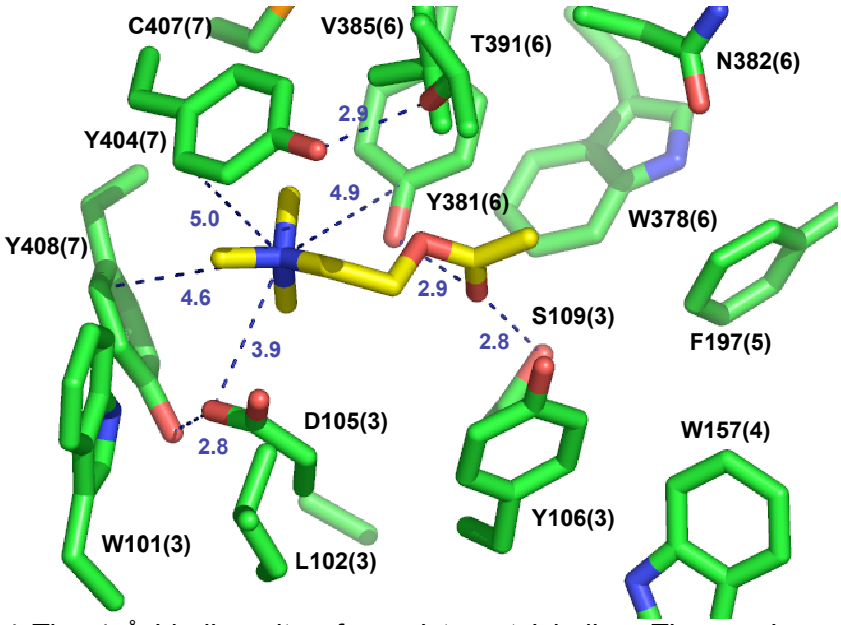


Fig. 14 The 4 Å binding site of agonist acetylcholine. The numbers shown in parentheses are the TM helix to which the residue belongs. Trp157 and Leu 102, which are within the 5.5 Å distance to acetylcholine, are also shown. Thr 391 is not within the acetylcholine binding site. We included Thr 391 in Fig. 10 to illustrate its inter-helical hydrogen bond with Tyr 404.

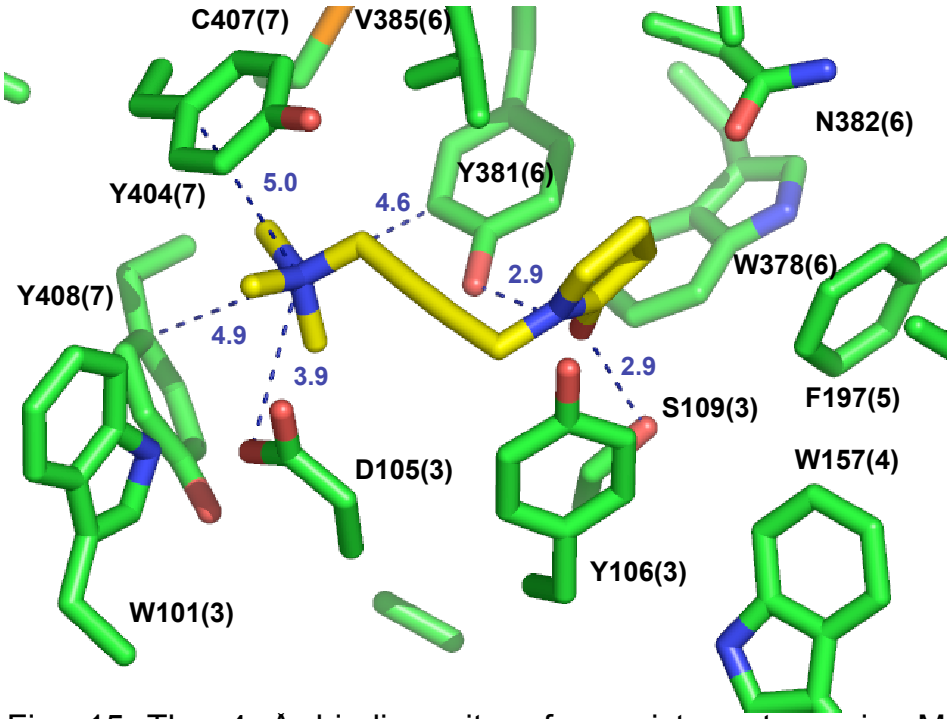


Fig. 15 The 4 Å binding site of agonist oxotremorine-M. The numbers shown in parentheses are the TM helix to which the residue belongs. Trp157, which is within 5 Å distance to oxotremorine-M, is also shown.

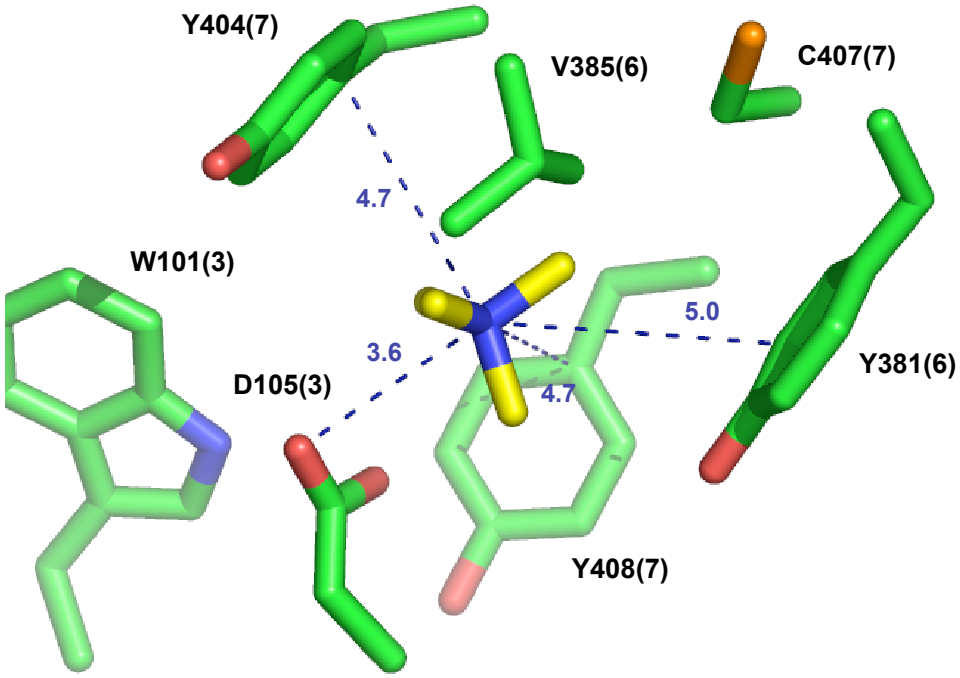


Fig. 16 The 4 Å binding site of agonist tetra-methylammonium. The numbers shown in parentheses are the TM helix to which the residue belongs.

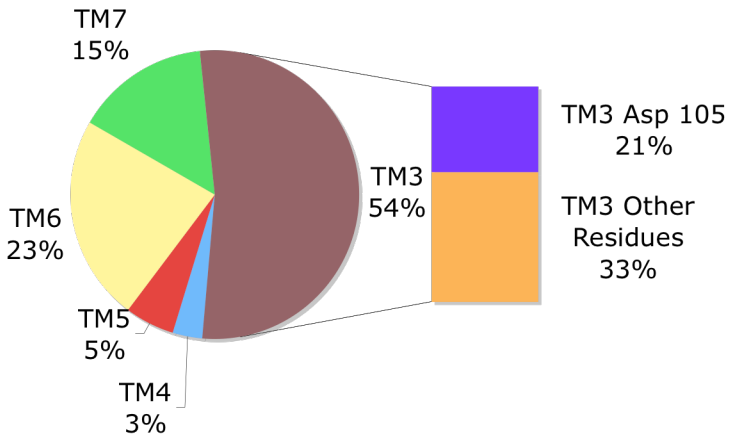


Fig. 17 Contribution of each TM region to the binding of the antagonists with quaternary ammonium head group. The total pair-wise interactions between the ligands and the residues within 6 Å binding sites were calculated for each TM. The energies were averaged from NMS and 4-DAMP binding sites.

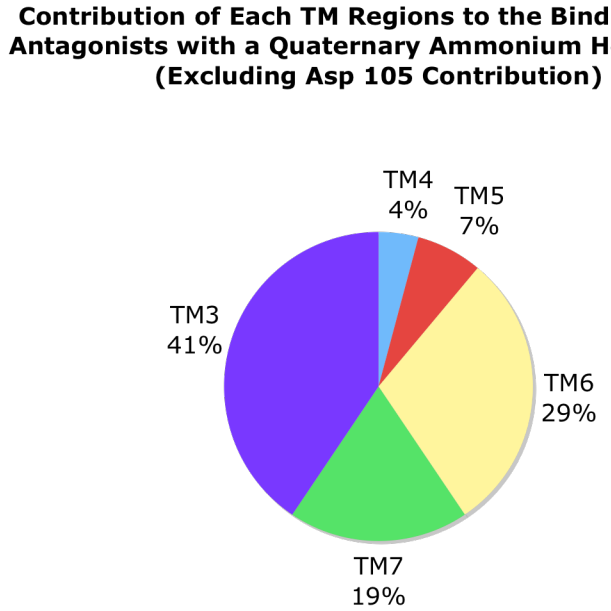


Fig. 18 Contribution of each TM region to the binding of the antagonists with quaternary ammonium head group excluding the Asp 105 contribution.

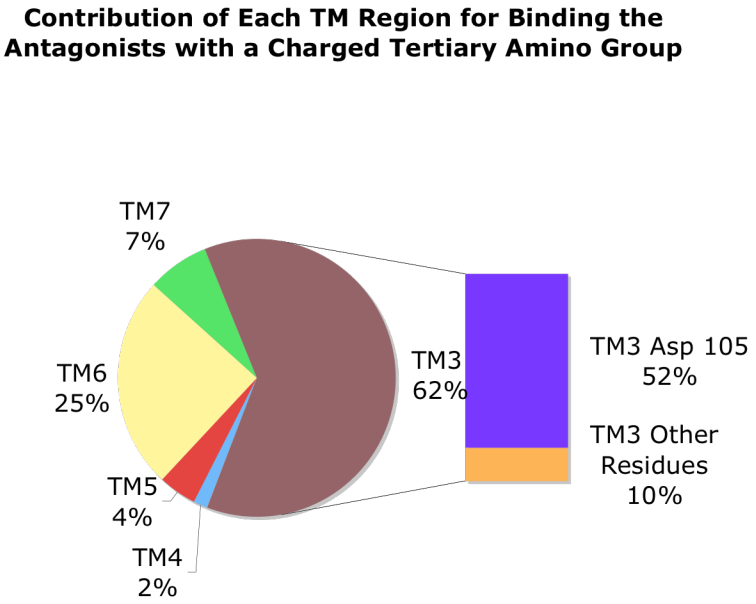


Fig. 19 Contribution of each TM region to the binding of the antagonists with a charged tertiary amino head group. The total pair-wise interactions between the ligands and the residues within 6 Å binding sites were calculated for each TM. The energies were averaged from QNB and pirenzepine binding sites. Note that Asp 105 contributes the most to binding energy and was found experimentally to reduce QNB binding by more than 6000 fold when mutated to Ala.

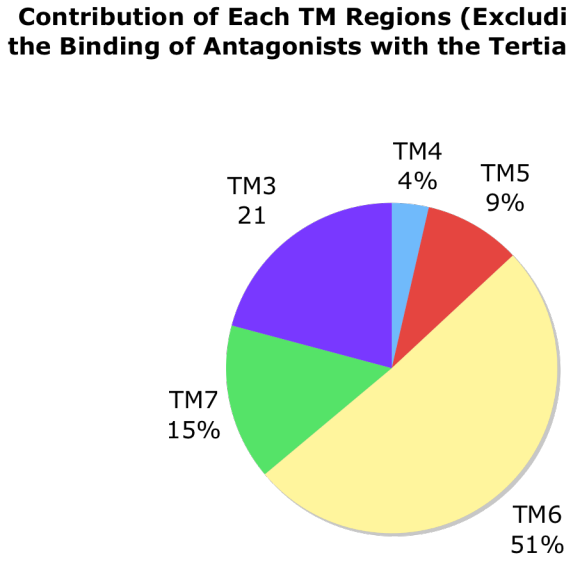


Fig. 20 Contribution of each TM region to the binding of the antagonists with tertiary amine head group excluding the Asp 105 contribution.

Contribution of Each TM Regions to Acetylcholine Binding

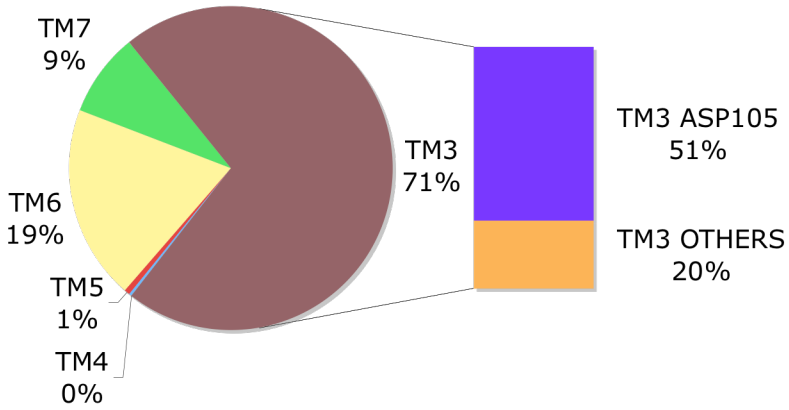


Fig. 21 Contribution of each TM region to the binding of the agonists. The total pair-wise interactions between the ligands and the residues within 6 Å binding sites were calculated for each TM.

Contribution of Each TM Regions to Acetylcholine Binding (Excluding Asp 105)

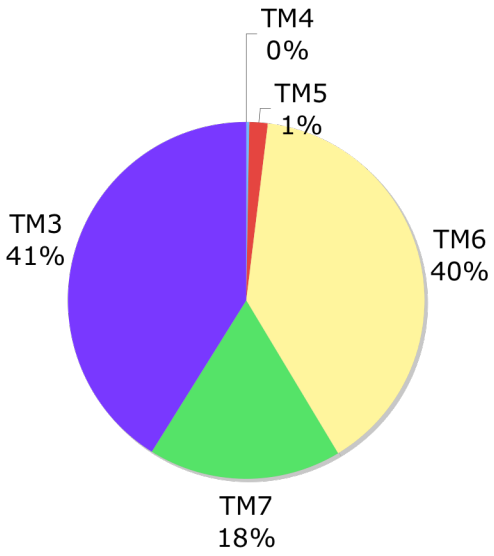
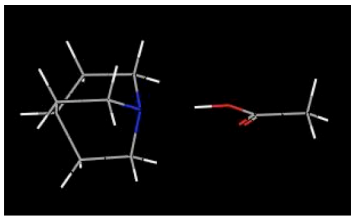
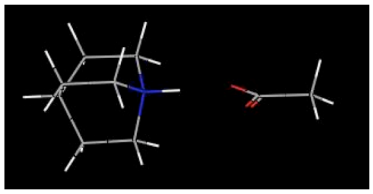


Fig. 22 Contribution of each TM region to the binding of the agonists excluding the Asp 105 contribution.

Environment	Dielectric Constant	Proton Transferred ?
Vacuum	1.0	Yes
Cyclohexane	2.0	Yes
Benzene	2.3	Yes
Methanol	33.6	No
Water	80.4	No



Proton Transferred



No Proton Transfer

Fig. 23 B3LYP/6-31G** geometry optimization for investigating the likelihood of proton transfer from QNB head group to Asp 105 on M1.

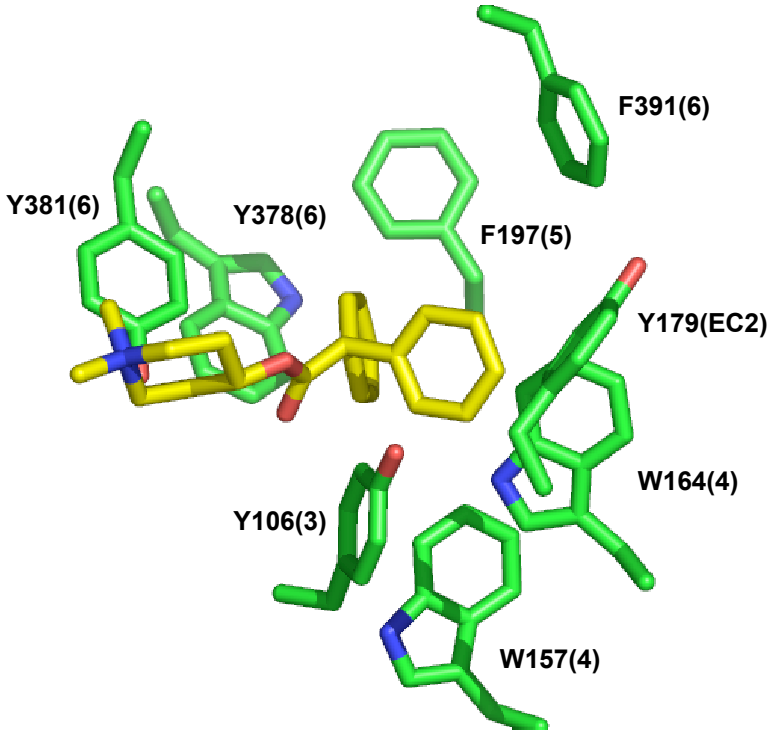


Fig. 24 The extensive aromatic network at the 4-DAMP binding site.

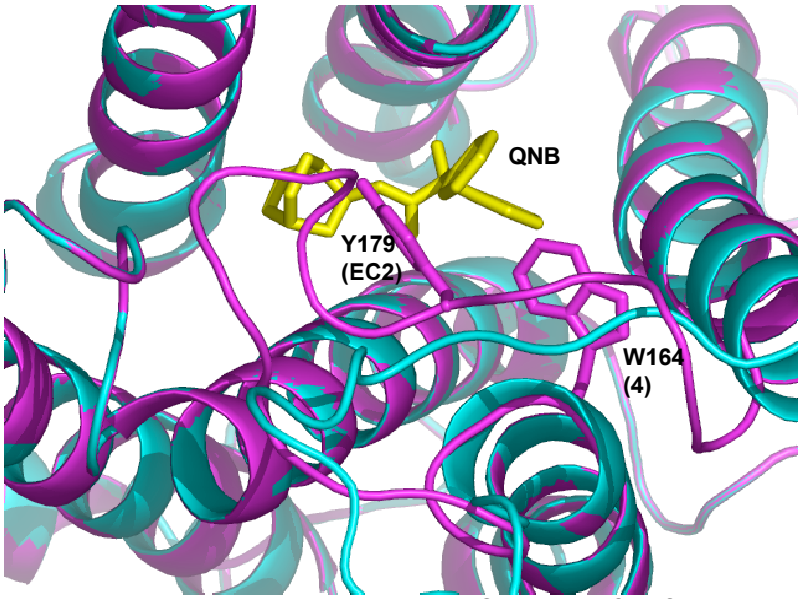


Fig. 25 The structure in purple has a closed EC 2 loop after further annealing dynamics, allowing the EC 2 residues, the ligand, and residues at the extra-cellular end of TM4 and TM5 to move, while keeping all the other residues fixed. The original structure (open loop) is shown in cyan. The closure of the EC2 loop brought Tyr 179 at the EC 2 loop and Trp 164 at the extra-cellular end of TM4 to within 5 Å of QNB. These aromatic residues extend the aromatic cage at the QNB binding site described earlier.

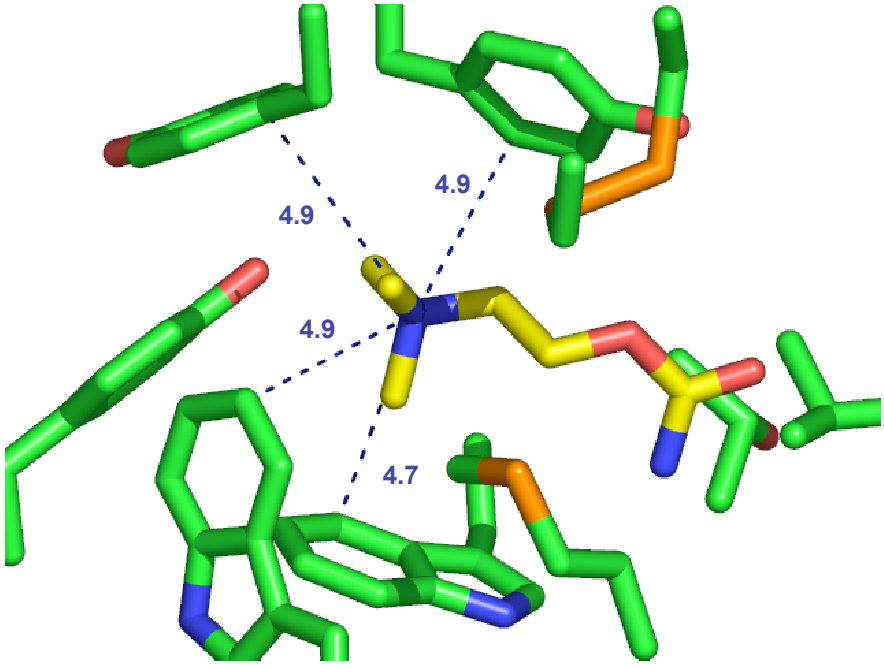


Fig. 26 Acetylcholine binding to the nicotinic acetylcholine receptor. The cation-π interaction distances that stabilize the acetylcholine head group are highlighted.

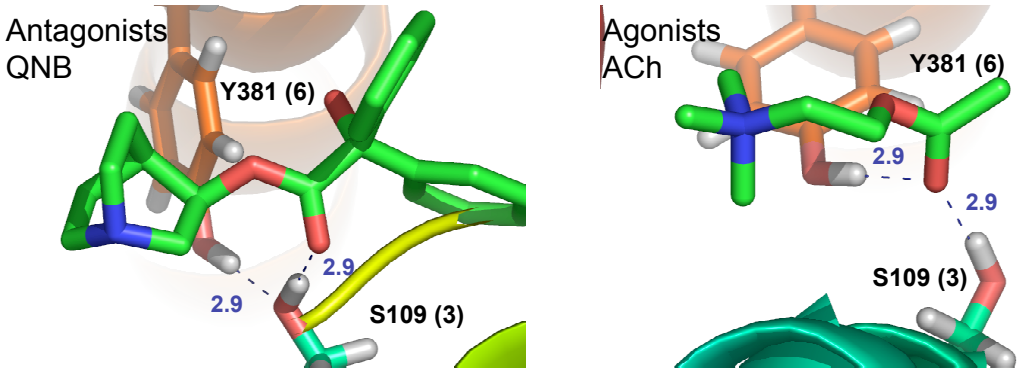


Fig.27 Comparison of the agonist and antagonist binding. The interhelical hydrogen bond between Tyr 381 on TM 6 and Ser 109 on TM3 is broken when agonists bind.

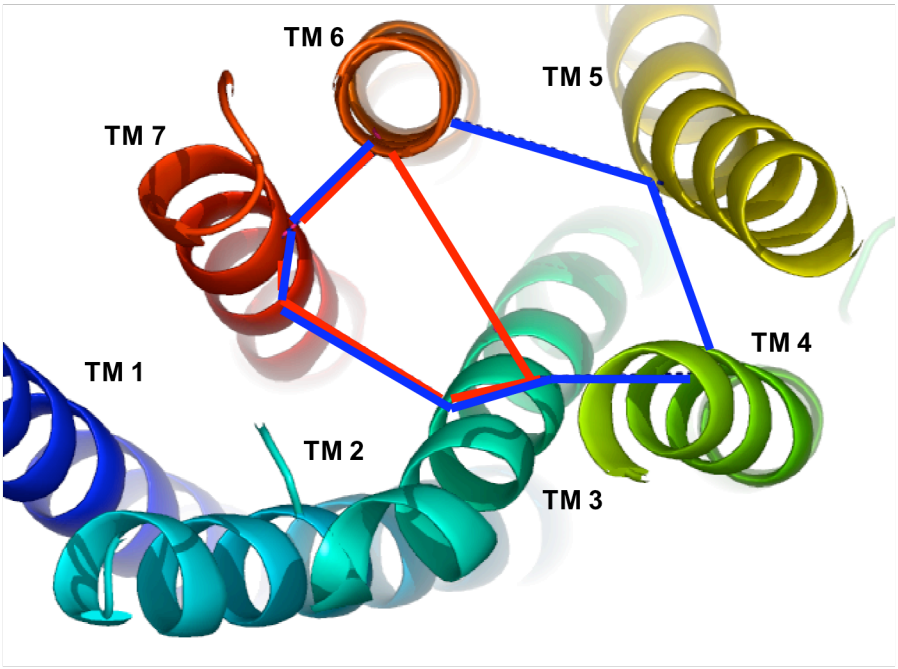
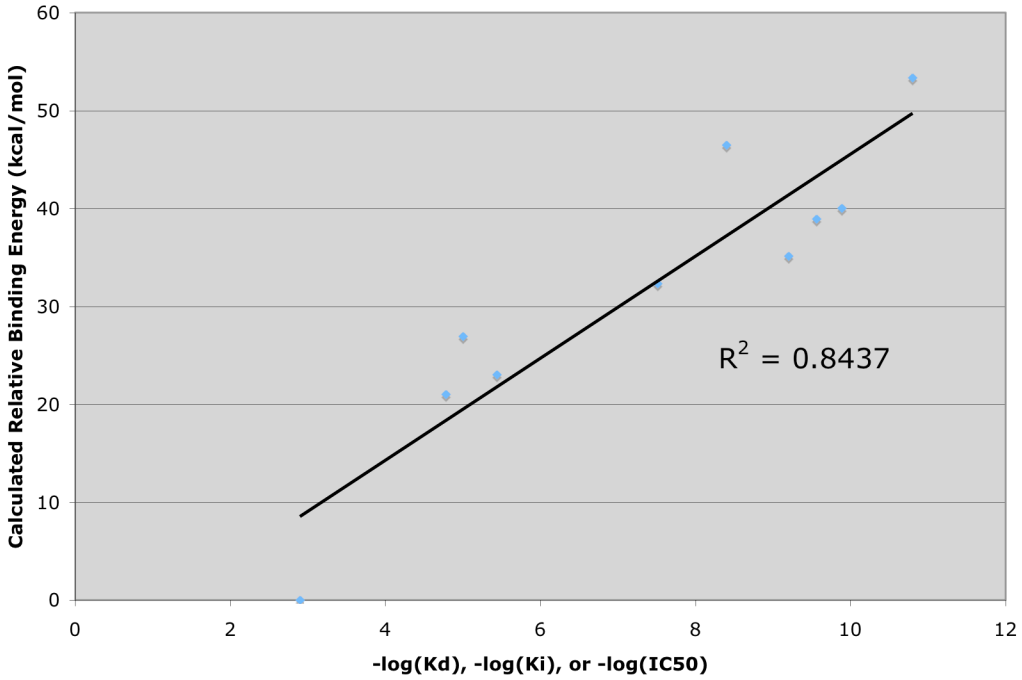


Fig. 28 Comparison of binding regions for antagonists (in blue) and agonists (in red). There is a significant overlap of the two binding regions in TM 3, 6, and 7, while antagonist regions extend to TM 4 and 5.



B

Fig. 29 Comparison of calculated binding energies all the 10 ligands used in this study with the experimental dissociation constants for the M1 receptor. Here a more positive energy indicates stronger binding. No corrections to the binding energies were made for zero-point energy, dynamics, or entropy.

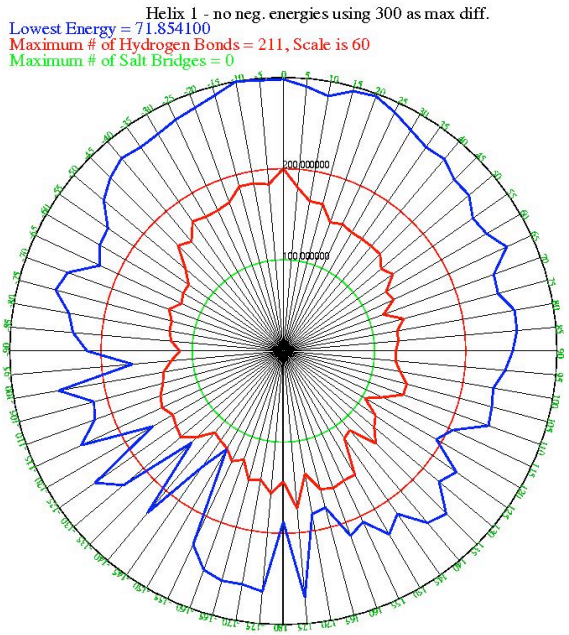


Fig. 30 Energetic rotational scan of TM1. The energy values (in blue) are scaled so that the best are at the outer circle. The energies become progressively worse closer to the center. Only the negative energy values are plotted. Therefore, if a particular rotational angle has a positive value in energy, it is plotted as zero (in the center of the plot). The red line represents the number of hydrogen bonds in the structure with the outer circle having the maximum number of hydrogen bonds. Both the maximum number of hydrogen bonds and lowest energy happen at degree zero for TM1, which corresponds to orientation in the final structure reported. There is no salt bridge between TM 1 to the other

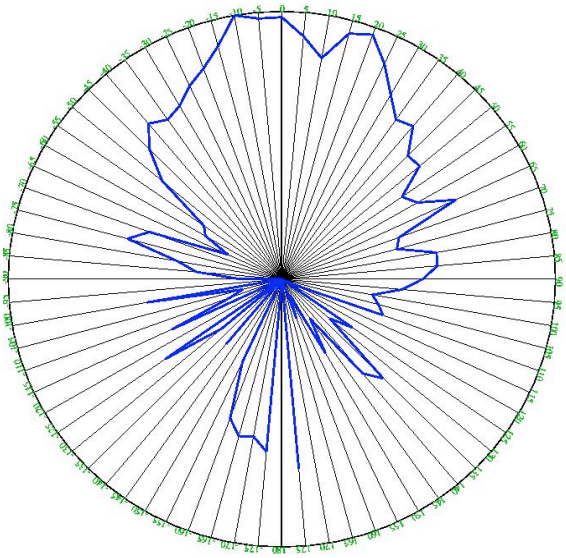


Fig. 31 Energetic rotational scans of TM1 with only the top 1/3 of the lowest energy values plotted to highlight the energy minimums. The energy values (in blue) are scaled so that the best are at the outer circle. The energies become progressively worse closer to the center.

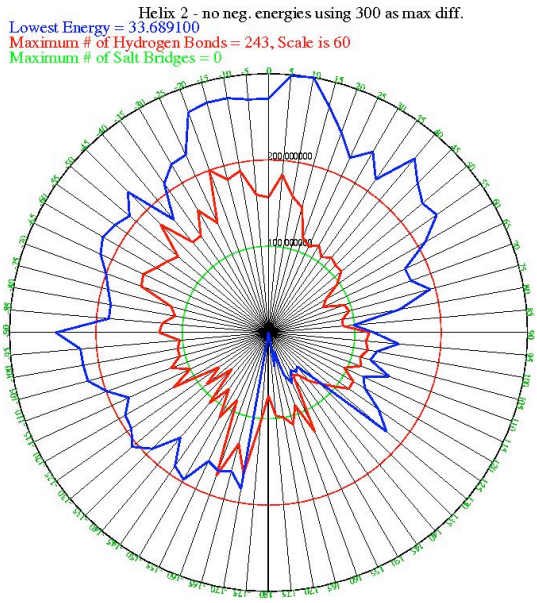


Fig. 32 Energetic rotational scans of TM2. The final structure corresponds to the zero degree rotational orientation.

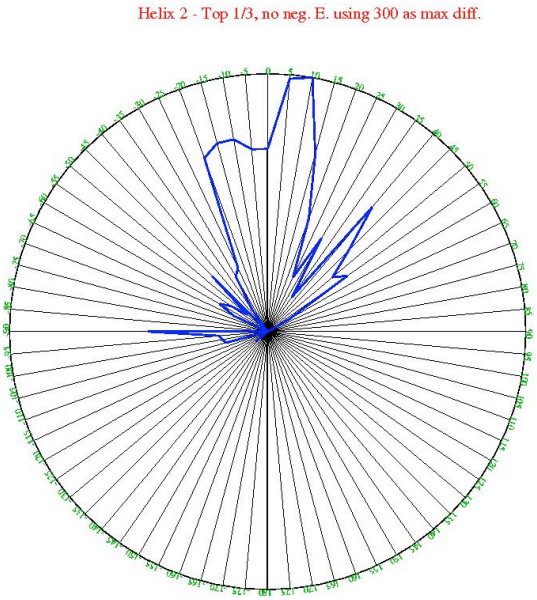


Fig. 33 Energetic rotational scans of TM2 with only the top 1/3 of the lowest energy values plotted to highlight the energy minimums.

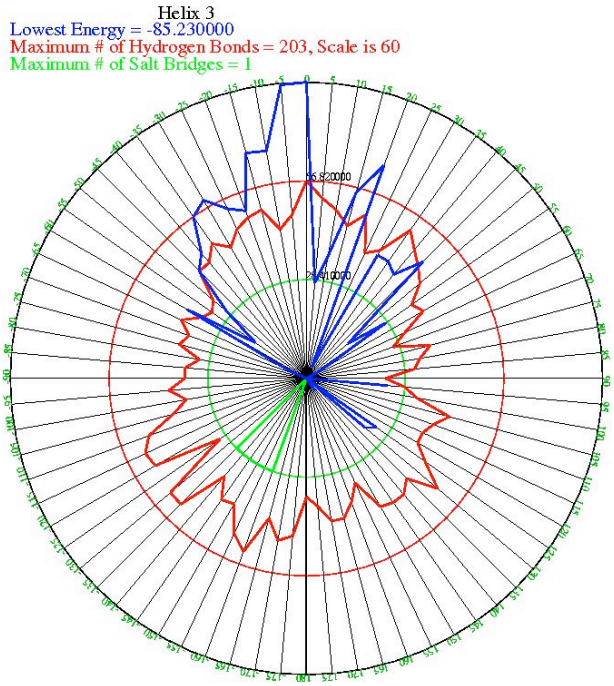


Fig. 34 Energetic rotational scans of TM3. Only the negative energy values are plotted. Therefore, if a particular rotational angle has a positive value in energy, it is plotted as zero (in the center of the plot). Both the 20 degree and the zero degree conformation were docked. The final structure corresponds to the 20 degree rotational orientation.

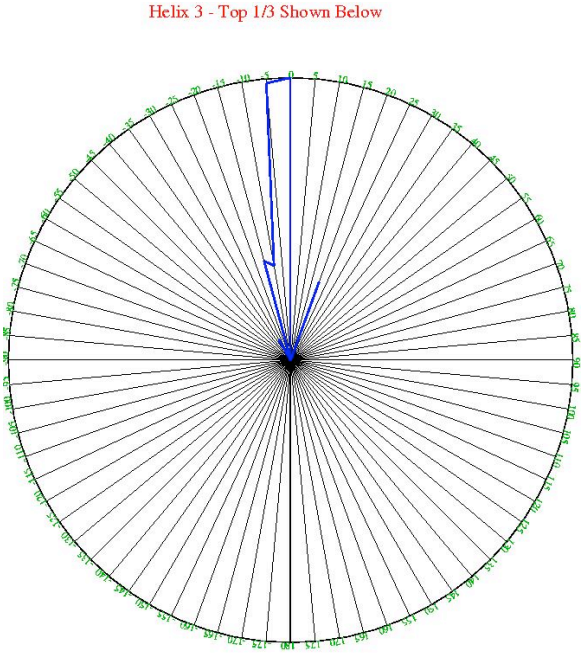


Fig. 35 Energetic rotational scans of TM3 with only the top 1/3 of the lowest energy values plotted to highlight the energy minimums.

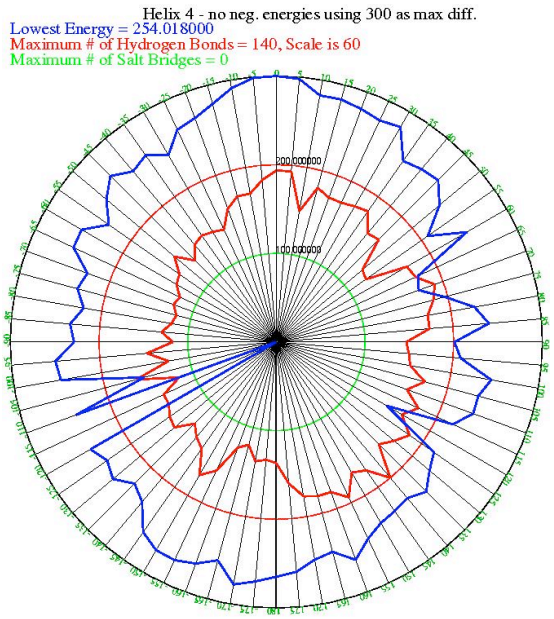


Fig. 36 Energetic rotational scans of TM4. The final structure corresponds to the zero degree rotational orientation.

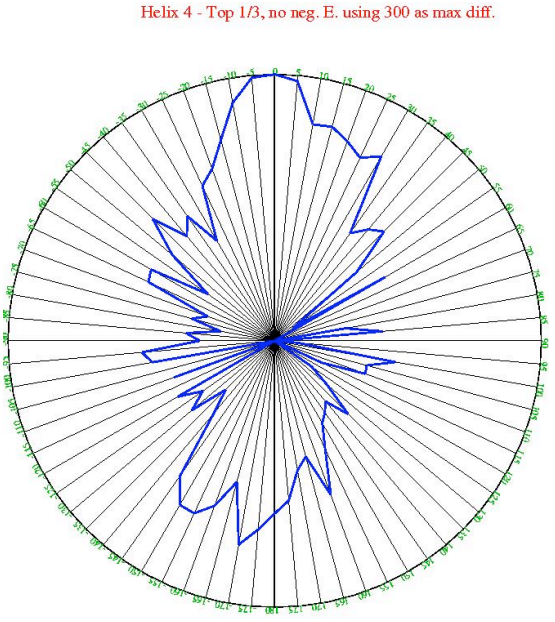


Fig. 37 Energetic rotational scans of TM4 with only the top 1/3 of the lowest energy values plotted to highlight the energy minimums.

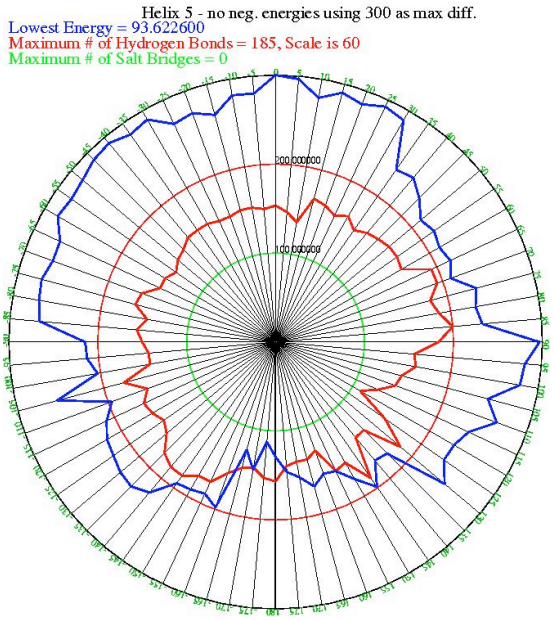


Fig. 38 Energetic rotational scans of TM5. The final structure corresponds to the zero degree rotational orientation.

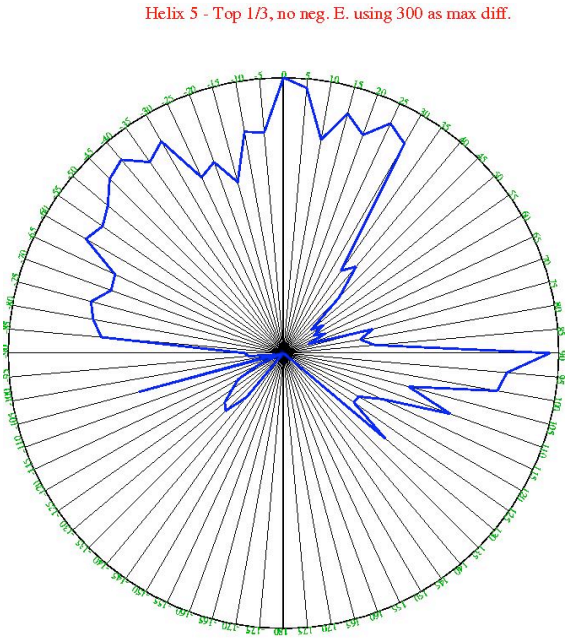


Fig. 39 Energetic rotational scans of TM5 with only the top 1/3 of the lowest energy values plotted to highlight the energy minimums. The two minimums (0 and 90 degree) seem to correspond to mutations that selectively affect antagonists binding (ground state) and agonists binding (activated) state, respectively,

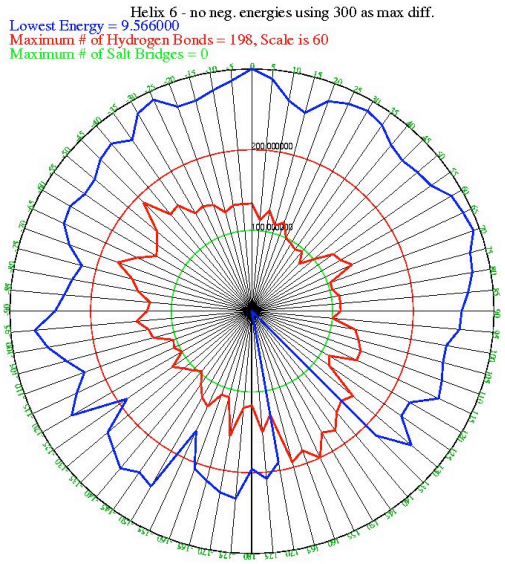


Fig. 40 Energetic rotational scans of TM6. The final structure corresponds to the zero degree rotational orientation. The energies of the rotational angles from 140-165 degrees are too large (due to bad van der Wall clashes) to be plotted.

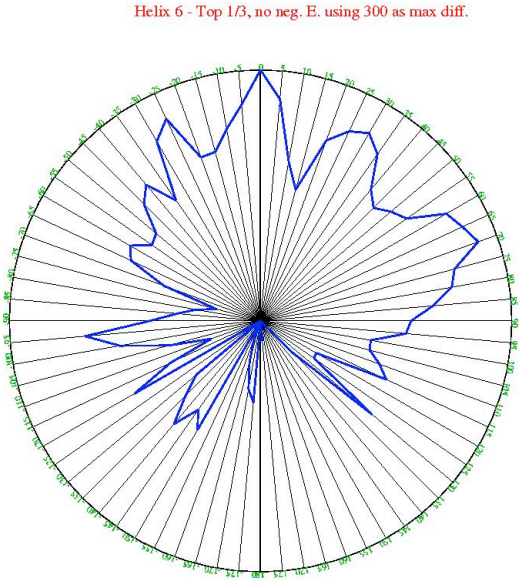


Fig. 41 Energetic rotational scans of TM6 with only the top 1/3 of the lowest energy values plotted to highlight the energy minimums.

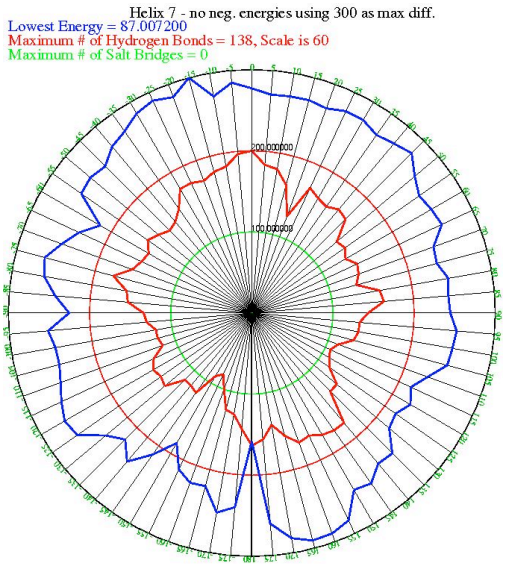


Fig. 42 Energetic rotational scans of TM7. The final structure corresponds to the zero degree rotational orientation.

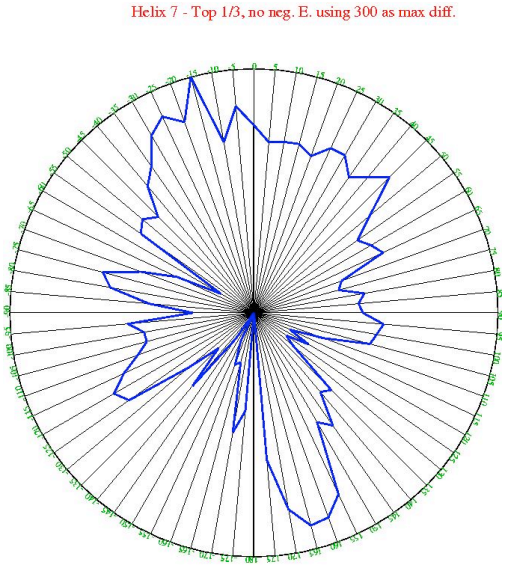


Fig. 43 Energetic rotational scans of TM7 with only the top 1/3 of the lowest energy values plotted to highlight the energy minimums.

Part 2: Cation-π Interactions at the Nicotinic Acetylcholine Receptor Binding Site: A Quantum Mechanical Study

Abstract

Cation-π interactions play a critical role in ligand binding and recognition in nicotinic acetylcholine receptors as well as many other biological systems. However, conventional force fields do not describe cation-π interactions accurately. We used high-level quantum mechanics with a high quality basis set to quantify the cation-π interactions in the environment of a protein using coordinates extracted from the crystal structure of carbamylcholine binding to Acetylcholine-binding Protein, a water-soluble homolog of the ligand binding domain of nicotinic receptors. Specifically, we extracted the carbamylcholine quaternary ammonium head group and the five aromatic sidechains interacting with it as our model system to calculate the binding energy contribution of each residue with X3LYP/B3LYP DFT methods. The results show that Tyr 192 and Trp 143 contribute the most (accounting for 31% and 29%, respectively) to the binding of the ligand head group, followed by Trp 53 (23%) and Tyr (17%), while Tyr 89 does not contribute in binding energy. In addition, we calculated the effect of fluorinated Trp 143 on binding and observed a decrease in binding energy with each additional fluorine substitution as it decreases the ability of Trp 143 to form cation-π interactions. The predicted change in binding energy correlates excellently with the experimental data in nicotinic acetylcholine receptor. Our results suggest that quantum mechanics can accurately predict cation-π bindings in a protein environment and provide a good model system in developing force fields to better describe cation-π interactions.

Chapter 4

CATION-π INTERACTIONS AT THE NICOTINIC ACETYLCHOLINE RECEPTOR BINDING SITE: A QUANTUM MECHANICAL STUDY

Introduction

The cation-π interaction is an important, general force for protein stability (1), protein folding (2), and molecular recognition in biological receptors, including both the nicotinic and muscarinic receptors (3). It is most strongly influenced by electrostatic forces, involving the interaction of a cation with the large, permanent quadrupole moment of an aromatic ring (1). Quantum mechanics were used to study the cation-π interactions between individual aromatic side chains and the cation-containing heterocyclic ring fragments of nicotinic ligands (1, 3). However, these pair-wise studies do not provide a holistic view of how the acetylcholine quaternary ammonium head group is stabilized in the environment of an aromatic cage, how many aromatic residues are needed, and how much each aromatic side-chain contributes to ligand binding.

Fortunately in 2004, the crystal structure of carbamylcholine binding to Acetylcholine-binding Protein (AChBP), a water-soluble homolog of the ligand binding domain of nicotinic receptors, became available (4). AChBP is found in the snail Lymnaea stagnali. It is produced and stored in glial cells, and released in an acetylcholine-dependent manner into the synaptic cleft, where it modulates synaptic transmission. AChBP is 210 residues long and forms a stable homopentamer. Nearly all residues conserved within the nicotinic acetylcholine receptor family are present in AChBP, including the ones that are relevant for ligand binding. Moreover, AChBP binds known agonists and antagonists (such as acetylcholine, nicotine, d-tubocurarine and α-bungarotoxin) of nicotinic receptors (5), so AChBP has been used extensively as a model system to understand the nicotinic receptor binding site. The availability of the carbamylcholine-AChBP complex crystal structure enabled us to examine, in the context of a protein, the cation-π interactions at the nicotinic receptor binding site with quantum mechanical methods.

In addition, we calculated the effect of unnatural amino acids substitutions that were incorporated into the nicotinic receptor binding site using the in vivo nonsense-suppression method (6). We found an excellent correlation between the calculated binding energies and the experimental EC 50 results. The knowledge we learned from this study is instrumental in the prediction and refinement of the M1 muscarinic receptor binding site. It also provides a good model system for developing force fields that can better describe cation-π interaction in proteins.

Computational Methods

Structure Preparation

The protein structure used in this study was extracted from the 2.5 Å crystal structure (4) of carbamylcholine binding to Acetylcholine-binding Protein (PDB code: 1UV6), a water-soluble homolog of the ligand binding domain of nicotinic receptors (Fig. 1). The coordinates extracted include the carbamylcholine quaternary ammonium head group and the five aromatic side chains interacting with the head group. The hydrogen atoms were added to the PDB file and the β carbon of each side chain was modeled as a methyl group. We then optimized the hydrogen positions using conjugate gradient minimization and TVN molecular dynamics while keeping all the non-hydrogen atoms fixed in order to preserve the original crystal structure. The final structure used in this study is shown in Fig. 2. There are a total of 103 atoms in the system.

Binding Energy Calculations with Quantum Mechanical Methods

We used high-level quantum mechanics, the density-functional theory (DFT) methods, and a high quality basis set, ccpv-tz(-f) and 6-31G**, to accurately calculate the binding energy contribution of each aromatic side chain. Two DFT methods were used: the standard Becke three parameter Lee-Yang-Parr (B3LYP) and an enhanced extended hybrid density functional method (X3LYP) recently developed by our group. This method has been shown to significantly improve the accuracy for hydrogen-bonded and van der Walls complexes

(7). The following section explains how we calculated the binding energy contribution of each aromatic side chain in the complex using Trp 143 as an example:

(1) BE (the whole complex) = PE (the whole complex) – PE (the five aromatic sidechains only) – PE (the ligand head group)

(2) BE (the whole complex with Trp 143 removed) = PE (the whole complex with Trp 143 removed) – PE (the four aromatic sidechains only, with Trp 143 removed) – PE (the ligand head group)

(3) BE (contribution from Trp 143) = BE (the whole complex) - BE (the whole complex with Trp 143 removed)

BE stands for binding energy and PE stands for potential energy. The energy calculations were conducted for each aromatic side chain without solvation. We also calculated the pair-wise interaction energy between the ligand head group and each aromatic side chain as follows:

BE (pair-wise) = PE (the ligand and sidechain pair) – PE (the ligand) – PE (the sidechain)

Unnatural Amino Acid Substitutions

We substituted Trp 143 (which corresponds to Trp 149 in nicotinic receptors) with a series of fluorinated substituents: 5-F-Trp; 5,7-F2-Trp; 5,6,7-F3-Trp; and 4,5,6,7-F4-Trp. We calculated the binding energy contributions of these fluorinated Trp substitutions and compared the results to experimental EC50 data (6).

Results and Discussions

Fig. 3 lists the calculated pair-wise binding energy between each individual aromatic sidechain and the ligand head group. Fig. 4 lists the binding energy contribution of each aromatic sidechain in the protein complex. Two DFP methods (X3LYP and B3LYP) and two different basis sets (cc-pvtz-(f) and 6-31g**) were used to calculate the binding energy. Fig. 3 and Fig. 4 show that these methods gave similar results, although X3LYP consistently gave a slightly larger binding energy than B3LYP did, given the same basis set. The X3LYP was shown in other systems to outperform B3LYP in describing van der Waals and hydrogen bond interactions (7). The better binding energy from X3LYP observed in our system reflects X3LYP's ability to better describe the attractive part of the van der Walls energy. Because cc-pvtz(-f) is a larger basis set than 6-31g**, we also consider the results from cc-pvtz(-f) more accurate than 6-31g**.

Fig. 5 compares the binding energy contribution of each aromatic sidechain in the complex and the pair-wise binding energy, both calculated by X3LYP/cc-pvtz(-f) methods. The two sets of results show a similar trend in the relative importance of each

aromatic residue in ligand binding, although the value of the binding energy is 1 to 2 kcal/mol smaller in the complex than in pair-wise calculations due to the interactions among the aromatic sidechains in the complex.

The results show that in the crystal structure complex, Tyr 192 and Trp 143 contribute the most (accounting for 31% and 29%, respectively) to binding the ligand head group, followed by Trp 53 (23%) and Tyr (17%), while Tyr 89 does not contribute (Fig. 6). The ligand-sidechain pair gave similar results in binding energy contributions: Trp 143 (28%), Tyr 192 (26%), Trp 53 (23%), Tyr 185 (19%), and Tyr 89 (4%) (Fig. 6).

Mutagenesis studies suggested that the hydroxy group of the Tyr residue equivalent to Tyr 89 in AChBP is essential for high affinity acetylcholine binding in nicotinic acetylcholine receptors (8). However, when we mutated Tyr 89 to Phe and recalculated its binding energy contribution, we observed that the Phe substitution actually increased the binding energy slightly, while experimentally it significantly reduced acetylcholine binding affinity (8, 9). Our result suggests that this effect did not result from the cation-π interactions between Tyr 89 and the acetylcholine head group. The hydroxy group of Tyr 89 is hydrogen bonded to a main chain carbonyl group in the AChBP-carbamylcholine crystal structure complex (4). Tyr 89 might influence ligand binding by affecting the aromatic packing at the nicotinic receptors binding site, and this hydrogen bond might be important in keeping Tyr 89 at its current conformation.

Because of the importance of Trp 143 in ligand binding and the availability of experimental data of unnatural amino acid substitutions in nicotinic acetylcholine receptors, we calculated, in the complex, the effect of fluorinated Trp 143 on ligand binding. We observed a decrease in binding energy with each additional fluorine substitution as it decreased the ability of Trp 143 to form cation-π interactions. Fig. 7 shows that the predicted change in binding energy in the complex correlates excellently (R square = 0.98) with the experimental EC 50 data in the nicotinic acetylcholine receptor (6). Our results suggest that quantum mechanics (B3LYP/6-31g** or higher) can accurately predict cation-π bindings in a protein environment. The results also suggest that four aromatic sidechains are sufficient to stabilize the acetylcholine head group in the nicotinic receptor's binding site because the 5th ring, Tyr 89, contributes very little to the binding energy.

Conclusion

This quantum mechanical study provides a quantitative description of each aromatic residue's contribution in stabilizing the acetylcholine head group in the nicotinic receptor binding site. The result correlates excellently with the experimental EC 50 data where unnatural amino acids were incorporated, suggesting that quantum mechanics can accurately predict cation-π bindings in a protein environment. We hope that our results will not only lead to a better understanding of cation-π interactions critical to acetylcholine binding but also provide a good model system for future development of force fields to better describe cation-π interactions in proteins.

References

1. Mecozzi, S., West. A. P., Jr., & Dougherty, D. A. (1996) *Proc. Natl. Acad. Sci.* **93**, 10566-10571.

2. Gromiha, M. M., & Suwa, M. (2005) *Int. J. Biol. Macromol.* **35**, 55-62.

3. Schmitt, J. D., Sharples, C. G., & Caldwell, W. S. (1999) *J. Med. Chem*. **42**, 3066-3074.

4. Celie, P. H., van Rossum-Fikkert, S. E., van Dijk, W. J., Brejc, K., Smit, A. B., & Sixma, T. K. (2004) *Neuron* **41**, 907-914.

5. rejc, K., van Dijk, W. J., Klaassen, R. V., Schuurmans, M., van Der Oost, J., Smit, A. B., & Sixma, T. K. (2001) *Nature* **411**, 269-276.

6. Zhong, W., Gallivan, J. P., Zhang, Y., Li, L., Lester, H. A., & Dougherty, D. A. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12088-12093.

7. Xu, X., & Goddard, W. A., III (2004) *Proc. Natl. Acad. Sci. USA* **101**,2673-2677.

8. Nowak, M. W., Kearney, P. C., Sampson, J. R., Saks, M. E., Labarca, C. G., Silverman, S. K., Zhong, W., Thorson, J., Abelson, J. N., Davidson, N., et al. (1995) *Science* **268**, 439-442.

9. Sine, S. M., Quiram, P., Papanikolaou, F., Kreienkamp, H. J., & Taylor, P. (1994) *J. Biol. Chem.* **269**, 8808-8816.

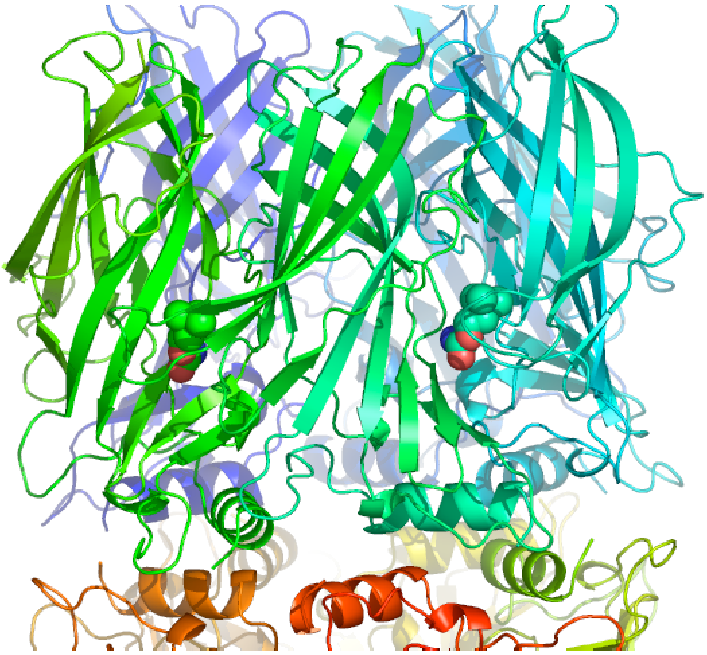
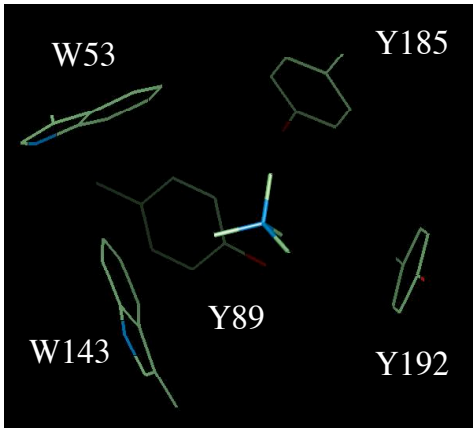


Fig. 1 The crystal structure of carbamylcholine binding to Acetylcholine-binding Protein.



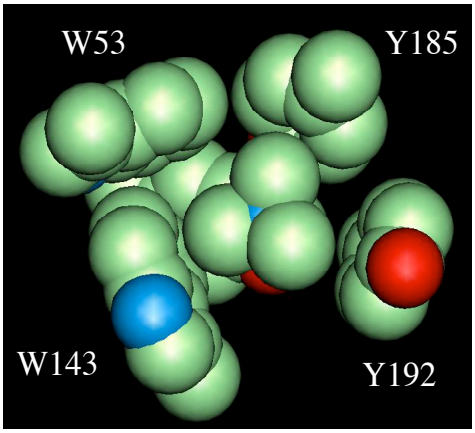


Fig. 2 The structure extracted as our model system. Hydrogen atoms are added to the structure.

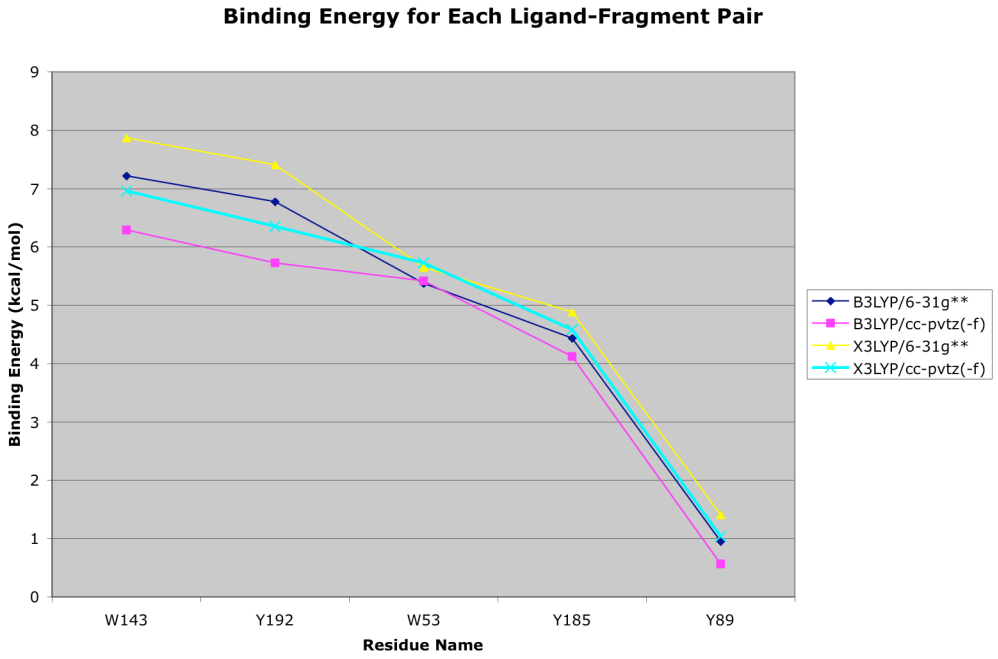


Fig. 3 The binding energy of each ligand-sidechain fragment pair.

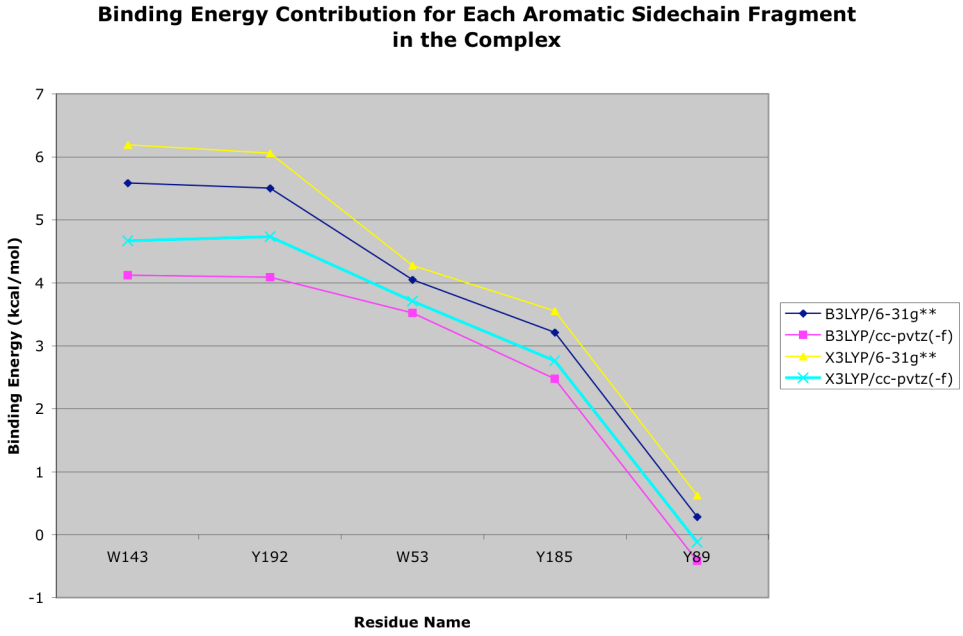


Fig. 4 The binding energy contribution of each aromatic side chain

Fig. 4 The binding energy contribution of each aromatic side chain fragment in the complex.

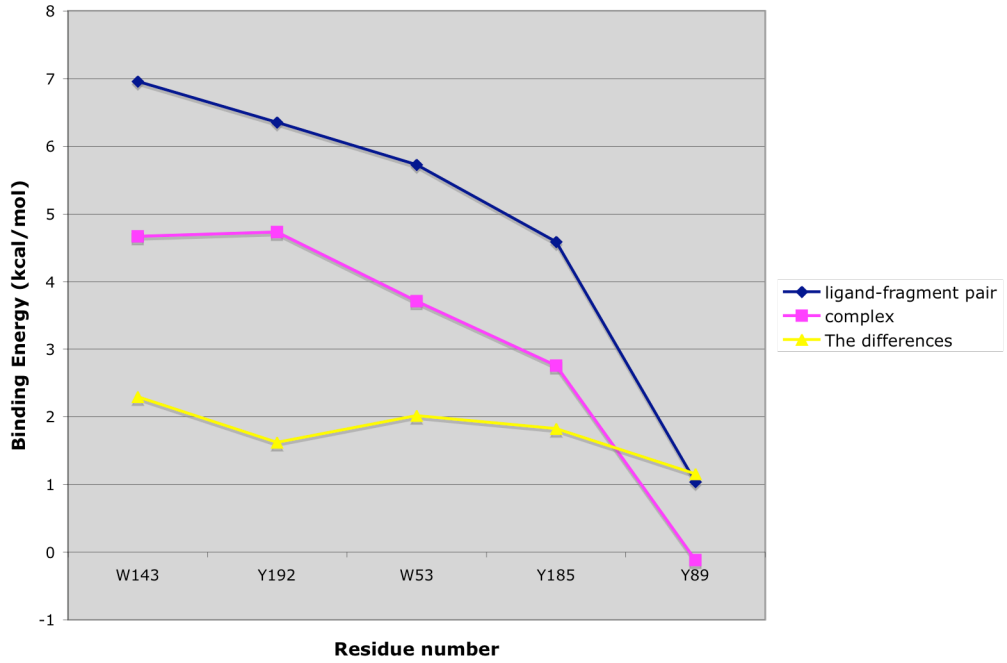


Fig. 5 Comparison of the binding energy contributions of each aromatic side chain fragment in the complex and in ligand-sidechain fragment pairs. The differences are due the interactions among the side chains in the crystal structure complex. The values plotted are results from X3LYP/cc-pvtz-(f) results.

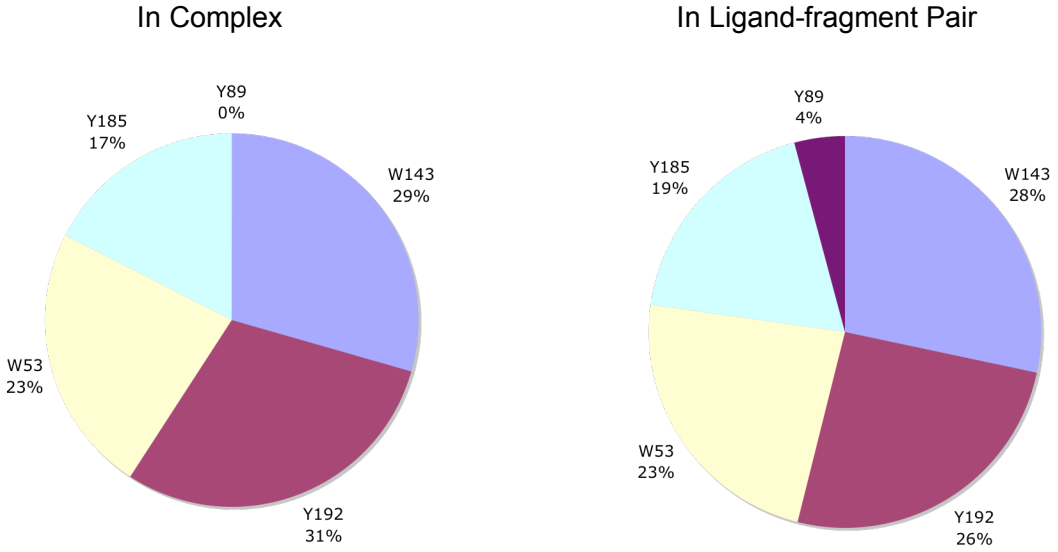


Fig. 6 The binding energy contribution of each aromatic side chain fragment in the complex and in ligand-side chain fragment pairs based on the X3LYP/ cc-pvtz(-f) calculation results.

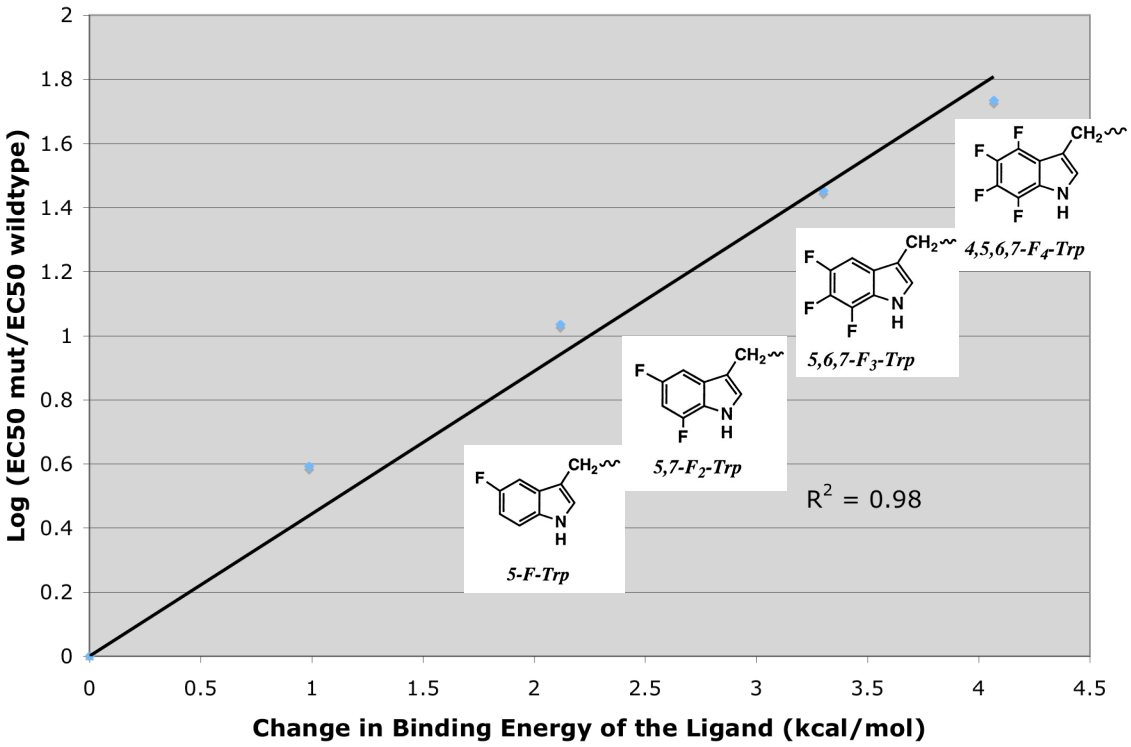


Fig. 7 Comparison of the changes in ligand binding energy from the widetype due to the fluorinated Trp substitutions (calculated by B3LYP/6-31g**) and the experimental EC50 values. The change (the amount of binding energy decreases) is shown in positive values where a larger number indicates a larger decrease in binding energy.

Part 3 Designing pH Responsive Proteins with Unnatural Histidines

Abstract

Histidine is known to modulate protein-protein interactions in a pH dependent manner. For example, the neonatal Fc receptor (FcRn) transports immunoglobulin G (IgG) across epithelia, binding to IgG in acidic vesicles (pH <6.5) and releasing IgG in the blood at pH 7.4. We designed a series of histidine derivatives by substituting the 2, 4, or both positions on the histidine imidazole ring with a number of functional groups (e.g. –CH3, F, Cl) that are small in size and lack the ability to form hydrogen bonds. Quantum mechanical calculations of the acid dissociation constants (pKa) show that these functional groups shift the histidine pKa upward or downward through electron donating or withdrawing capabilities. We report a list of histidine derivatives and their corresponding pKa values that can be used in various applications including drug design for tumor specificity (e.g. HER2-Herceptin antibody), drug delivery through pH sensitive hydrogels, drug recycling, catalysis, and biosensors development. An example of how these unnatural histidines can be used is illustrated with 2-methyl histidine incorporated in a c-Myc-Max heterodimer.

Chapter 5

INTRODUCTION TO THE DESIGN CONCEPT

Monoclonal antibodies have been used extensively in cancer therapy. Although these monoclonal antibodies are generally very specific for their targets, they tend to elicit certain side effects by binding to non-tumor tissue, often because their targets are not specifically expressed on tumor cells, but are also expressed (although in smaller numbers) on some healthy cells. Such side effects are generally undesirable, and there is a need for antibodies with an improved specificity. One such example is the monoclonal antibody Herceptin (Trastuzumab) against breast cancer (1). Herceptin binds to the human epidermal growth factor Type II receptor (HER-2/neu, the c-erbB2 gene product). It has a serious side effect that damages the hearts of patients, partly due to the expression of HER-2/neu receptors on both tumor cells and normal heart cells.

The pH of human blood is highly regulated and maintained in the range of 7.4-7.8, whereas the extracellular environment of tumor cells is more acidic (pH of 6.3-6.5) because of the accumulation of metabolic acids that are inefficiently cleared due to poor tumor vascularization (2). If the interaction between a tumor antigen and its therapeutic antibody can be modulated such that at low pH, the binding is favored, the tumor-antibody may have an added specificity for those tumor antigens, even though the same tumor antigens are also found on normal tissues.

Histidine is known to modulate protein-protein interactions in a pH dependent manner. For example, the neonatal Fc receptor (FcRn) transports immunoglobulin G (IgG) across epithelia, binding to IgG in acidic vesicles (pH <6.5) and releasing IgG in the blood at pH 7.4. FcRn binds to Fc with high affinity at lower pH when the Fc histidines 310, 435, 436 are positively charged, and releases Fc upon their deprotonation at higher pH (3). This concept was successfully used in rational drug design to increase the lifetime and potency of granulocyte colony-stimulating factor (GCSF) by mutating selected residues to histidines (4). The design of the GCSF mutant was based on the principle that neutral histidines could maintain relatively tight binding on the cell surface (pH > 7), but protonated histidines could lead to weaker binding in endosomal compartments (pH between 5 and 6), which results in increased recycling to the cell surface (4).

Protein engineering is a powerful tool for modification of the structural, catalytic and binding properties of proteins. The incorporation of unnatural amino acids can greatly extend the scope and impact of protein engineering methods. Unnatural amino acids can be incorporated into proteins by various methods including chemical synthesis, nonsense suppression methods (5), or by utilizing a modified aminoacyl-tRNA synthetase (6). We designed a series of unnatural histidines that can be incorporated into proteins for pH-sensitive binding. It is known in organic chemistry that the strengths of a series of substituted benzoic acids are different from the strength of benzoic acid itself. The extent of the effect is correlated with the electron withdrawing or donating capabilities of the substituents (7). We applied this concept in designing a series of histidine derivatives by

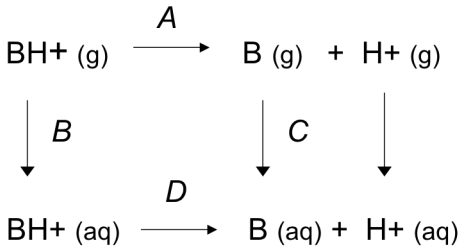
substituting the 2, 4, or both positions on the histidine imidazole ring with a number of functional groups (e.g. –CH3, F, Cl) that are small in size and lack the ability to form hydrogen bonds. Quantum mechanical calculations of the acid dissociation constants (pKa) show that these functional groups shift the values of histidine pKa upward or downward based on their electron donating or withdrawing capabilities. In the next chapter, we report a list of unnatural histidines and their corresponding pKa values ranging from -8.4 to 8.2 (an enormous difference of more than 10^{16}), with 18 different values in between. These unnatural histidines and their combinations can be used in a wide range of applications including drug design for tumor specificity as we described above, drug delivery through pH sensitive hydrogels, drug recycling, catalysis, and biosensors development. An exemplary application of these histidine analogs is illustrated with 2-methyl histidine, which shifts the pKa value up for about 1 unit, incorporated in a c-Myc-Max heterodimer.

Chapter 6

ACID DISSOCIATION CONSTANTS OF UNNATURAL HISTIDINES AND AN EXEMPLARY APPLICATION

The Acid Dissociation Constants of Imidazole Derivatives

We conducted quantum mechanical calculations to derive the acid dissociation constants (pKa) of a series of imidazole derivatives using the Schrödinger's pKa prediction module in Jaguar™ 5.5 software. The detailed algorithm is described in the Jaguar User Manual (8) and summarized below. The following thermodynamic cycle was used to calculate the pKa value of a molecule in aqueous solution.



The pKa is related to *D* by

$$pKa = (1 / 2.3\ R\ T)\ D \quad \text{and}$$

$$A = \Delta H - T\,\Delta S = E_{B\,(g)} - E_{BH+\,(g)} + 5 / 2\ R\ T - T\,\Delta S$$

The algorithm calculated parts *A, B*, and *C* of the above cycle and *D* can be obtained by

summing the free energy changes for these three components and the experimental value of –259.5 kcal/mol for the solvation free energy change of a proton.

To investigate the accuracy of our methods in predicting the pKa values of imidazole variants, we calculated the pKa of a series of imidazole derivatives listed in Fig. 2 and compared our results to the experimental pKa values (9, 10). Fig. 1 shows the imidazole structure and lists the positions of substitutions on the imidazole referred to in Fig. 2. Fig. 2 shows that our methods are very accurate in predicting the experimental pKa values of these imidazole derivatives and suggests that no further adjustments of the calculated results are needed to accurately predict the experimental pKa values (9, 10).

We then substituted the hydrogens at the 2, 4, or both positions of the imidazole ring with the functional groups listed in Table 1. These functional groups were selected because they are relatively small in size and lack the ability to form strong hydrogen bonds that could affect protein binding. Our calculations suggest that these groups shift the imidazole pKa upward or downward depending on their electron donating or withdrawing capabilities. Table 1 lists the calculated pKa values of these imidazole derivatives. It is apparent that, even with the limited choice of just 6 types of small functional groups (e.g. -CN, -F, -Cl, $-CH_2F$, $-OCH_3$, or $-CH_3$), and two potential substitution positions (position 2 or 4 on the imidazole ring), the pKa of the resulting imidazole derivatives can range from a low of -7.1 to a high of 8.5, with 18 different values in between. A combination of these functional groups is expected to create more

pKa values.

The Additive Effect of 2, 4 Substitutions

Table 1 shows that for a given functional group, the pKa change in 2, 4 double substitutions are approximately the sum of the pKa changes by 2 and 4 single substitutions. For example, 2-CN-imidazole changed the imidazole pKa 7.1 units (from 6.9 to -0.2) and 4-CN-imidazole changed the pKa 6.6 units (from 6.9 to 0.3). The pKa of 2, 4-CN-imidazole is -7.1, which is a 14 units change and approximately equal to the changes of 2-CN and 4-CN substitutions combined. We observed similar trends for all other functional groups studied, suggesting that the electronic withdrawing or donating effect of 2, 4 substitutions are additive (Table 1). This result suggests that we can create more pKa values by substituting different functional groups at the 2, 4 positions of imidazole and we can estimate these pKa values from the single substitution results. This additive effect was verified by the published experimental pKa of various imidazole derivatives (9, 10). For example, the pKa changes (ΔpKa) in 2, 4-di-CH3 and 2, 4-di-nitro are approximately equal to ΔpKa of the 2 and 4 single substitution combined. This further validates the accuracy of our calculations.

The Acid Dissociation Constants of Histidine Analogs

We conducted similar quantum mechanical calculations to derive the pKa of a series of histidine derivatives. The structure of the histidine molecule is shown in Fig. 3.

We extended the histidine molecule to include the adjacent two α-carbon atoms in the protein backbone. The backbone was built in an α-helical conformation (which corresponds to the conformation in a leucine zipper). The structure was minimized with implicit solvent described by the surface Generalized Born (GB) continuum solvent model (11) before the quantum mechanical pKa calculations.

We substituted the hydrogens at the 2, 4, or both positions of the histidine imidazole ring with the same set of functional groups as the ones used in imidazole. As we mentioned earlier, these functional groups were selected because they are relatively small in size and lack the ability to form strong hydrogen bonds that could affect protein binding. The pKa was calculated for position 3 of the imidazole ring (Fig. 1 and Fig. 3). Our calculations suggest that these groups shift the histidine pKa (6.0 in experimental measurement and 5.8 in our calculation) upward or downward depending on their electron donating or withdrawing capabilities. Table 2 lists the calculated pKa values of the unnatural histidines that can be used in various applications to control pH-responsive protein binding. It is apparent that, even with the limited choice of just 6 types of small side-chain groups (e.g. -CN, -F, -Cl, $-CH_2F$, $-OCH_3$, or $-CH_3$), and two potential substitution positions (position 2 or 4 on the imidazole ring) of a single natural amino acid (e.g. histidine), the side-chain pKa of the resulting histidine analog can range from a low of -8.4 to a high of 8.2 (an enormous difference of more than 10^{16}), with 18 different values in between. A combination of different small side-chain groups at different ring positions is expected to create more pKa values.

Compared with a single value (pKa = 5.8) of the natural histidine residue, this

range greatly expanded the possibility of modulating pH-sensitive binding of a protein bearing such an unnatural amino acid.

Comparison of ΔpKa in His and Imidazole Derivatives

Table 3 compares the ΔpKa in His (pKa = 5.8) and imidazole (pKa = 6.9) and shows that the ΔpKa for each functional group is very similar in His and imidazole. This suggests that the same substitution might change the pKa for a similar magnitude in a protein, too, allowing us to estimate the ΔpKa of a specific substitution in a protein using Table 3.

The only major difference of ΔpKa in imidazole and His is observed in 4-OCH3 substitutions. Detailed examination of the quantum mechanical geometry optimization results (the first step in pKa calculation in Jaguar) revealed that in 4-OCH3-His, the OCH3 functional group was prevented from moving to the same, lower energy, conformation as in imidazole because of the steric hindrance of the protein backbone. The OCH3 functional group, unlike other functional groups selected, has an additional torsion angle. The conformational flexibility of OCH3 might affect the pKa value when incorporated into a protein and should be used with caution.

Incorporating 2-Methyl Histidine in the c-Myc-Max Heterodimer

Next, we substituted the histidine at the center of the c-Myc-Max heterodimeric leucine zipper (at position 8d of the helical wheel on Max in Fig. 4) with an unnatural histidine where its hydrogen atom at the 2 position of the imidazole ring was substituted with a methyl group. After we incorporated the 2-methyl histidine, we sampled the histidine rotamer library to calculate the energy of each conformation. Each structure was independently optimized with energy minimization in implicit solvent using the surface Generalized Born continuum solvent model. We also added counter ions to neutralize the system. The charges used for these histidine analogs were first derived from quantum mechanics calculations and then fitted to the CHARMM charges, so we can describe the backbone of all the residues in the leucine zipper with the same set of charges. We found that the lowest energy rotamer for 2-methyl histidine has the same conformation as the wild-type histidine in the NMR structure (12). The quantum mechanical pKa calculations (Table 3) suggest that this substitution will shift the histidine pKa (7.19 ± 0.03 experimental value (13)) up for about 1.1-1.4 units in the particular context of the leucine zipper.

Fig. 5 shows the positions of the wildtype and the 2-methyl histidines and their surrounding residues (within 5 Å) on c-Myc, including Glu 5, Leu 8, Ile 9, and Glu 12. To understand quantitatively how 2-methyl histidine interacts with its surrounding residues on c-Myc compared to the wildtype histidine, we calculated the interaction

energy between the histidine (in both its protonated and unprotonated form) and these residues. Our analysis, summarized in Table 4, suggests that the 2-methyl histidine and the wildtype histidine have very similar effects on target protein binding in the zipper. However, the effect was achieved at different pH values due to the side-chain pKa differences, which illustrates and conforms with our design purposes.

Conclusion

In this study, we introduced a novel concept of using unnatural histidines to modulate pH dependent binding. We designed a series of histidine analogs and reported their calculated pKa values, ranging from -8.4 to 8.2 (an enormous difference of more than 10^{16}). We illustrated the effect of incorporating one of these molecules, the 2-methyl histidine, into a leucine zipper c-Myc-Max heterodimer, which can be tested experimentally. As there are almost 100 industry related R & D projects utilizing conjugated antibodies for cancer therapy, we believe that the concept and the example described in this study will be very useful in designing drugs with improved tumor specificity, and in additional applications such as drug delivery, drug recycling, catalysis, and biosensors development.

References

1. Cho, H. S., Mason, K., Ramyar, K. X., Stanley, A. M., Gabelli, S. B., Denney, D. W., Jr., & Leahy, D. J. (2003) *Nature* **21**, 756-760.

2. Raghunand, N., Gatenby, R. A., & Gillies, R. J. (2003) *Br. J. Radiol.* **76,** S11-S22.

3. Martin, W. L., West, A. P., Jr, Gan, L., Bjorkman, P. J. (2001) *Mol. Cell.* **7**, 867-877.

4. Sarkar, C. A., Lowenhaupt, K., Horan, T., Boone, T. C., Tidor, B., & Lauffenburger, D. A. (2002) *Nat Biotechnol.* **20**,908-913.

5. Dougherty, D. A. (2000) *Curr Opin Chem Biol.* **4**, 645-652.

6. Hohsaka, T., & Sisido, M. (2002) *Curr Opin Chem Biol*. **6**, 809-815.

7. Vollhardt, K. P. C., & Schore, N. E. (1999) Organic Chemistry, Third Edition.

8. Jaguar 5.5 User Manual (2004) Schrodinger, L. L. C.

9. Perrin, D. D. (1965) Dissociation constants of organic bases in aqueous solution.

10. Kortüm, W. Vogel, and K. Andrussow (1962) Dissociation constants of organic acids in aqueous solution.

11. Bashford, D., & Case, D. A. (2000) *Annu. Rev. Phys. Chem.* **51**,129-152.

12. Lavigne, P., Crump, M. P., Gagne, S. M., Hodges, R. S., Kay, C. M., & Sykes, B. D. (1998) *J Mol Biol.* **281**, 165-81.

13. Lavigne, P., Kondejewski, L. H., Houston, M. E., Jr., Sonnichsen, F. D., Lix, B., Skyes, B. .D, Hodges, R. S., & Kay, C. (1995) *J. Mol. Biol.* **254**, 50

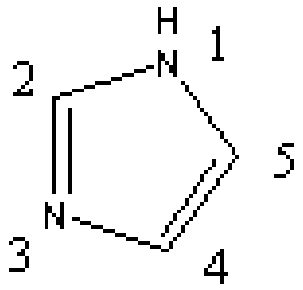


Fig. 1 The structure of the imidazole ring.

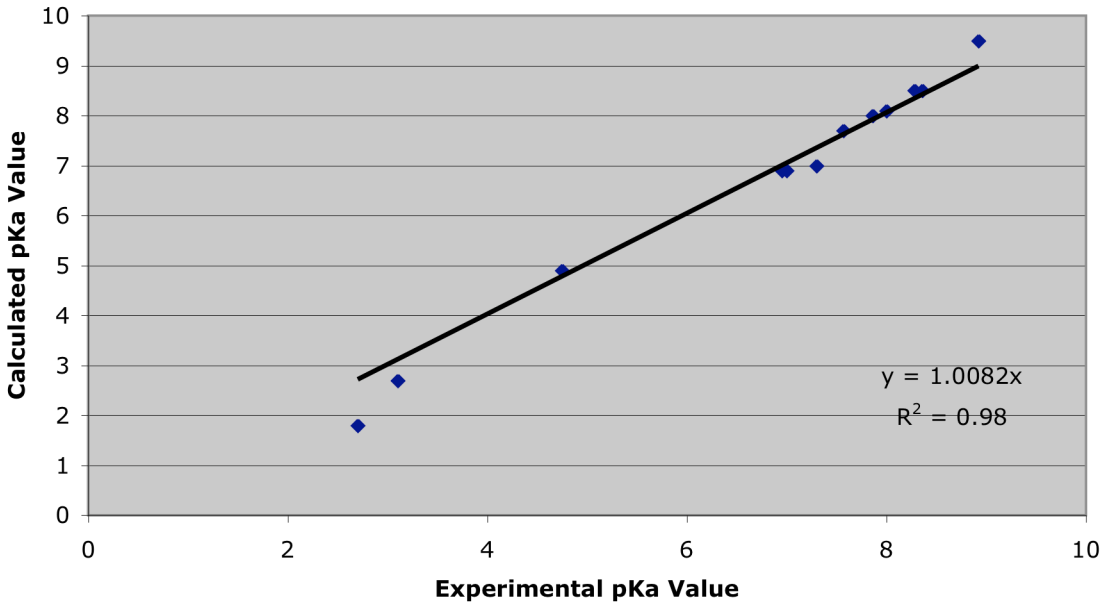


Fig. 2 Comparison of the calculated pKa values and the experimental pKa values. The numbering of the position of substitutions on the imidazole ring corresponds to what was shown in Fig. 1. The data points (from left to right) correspond to substitutions of 2-FI, 4-Cl-1-CH3, 5-Cl-1-CH3, wildtype imidazole, 1-CH3, 1-C2H5, 4-CH3, 2-CH3, 2-C2H5, 2,4-diCH3, 4, 5-diCH3, 2, 4, 5-triCH3 imidazoles.

	Position 2	Position 4	Position 2,4
CN	-0.2	0.3	-7.1
F	1.8	1.7	-3.4
Cl	2.2	2.8	-1.6
OCH3	5.3	6.0	5.0
CH2F	5.3	4.8	3.5
H	6.9	6.9	6.9
CH3	8	7.6	8.5

Table. 1 The calculated pKa values of the imidazole derivatives where the 2, 4 or both positions of the imidazole ring are substituted with the functional groups listed.

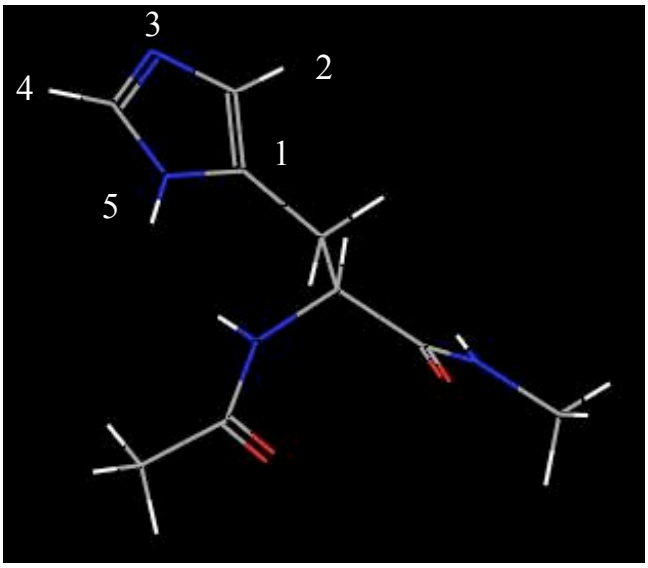


Fig. 3 The structure of the histidine. The numbers listed are the positions of substitutions on the imidazole ring. Position 3 is the protonation site where the pKa value was calculated.

His	Position 2	Position 4	Position 2,4
CN	-1.2	-0.6	-8.4
F	0.9	0.4	-4.1
Cl	1.4	1.2	-2.8
CH2F	4.1	4.4	2.4
OCH3	4.9	3.5	3.5
H	5.8 *	5.8	5.8
CH3	7.2	6.9	8.2

Table. 2 The calculated pKa values of histidine analogs (with the geometry shown in Fig. 3) where the 2, 4 or both positions of its imidazole ring are substituted with the functional groups listed. The free His pKa experimental value is 6.0. The calculated value is 5.8.

	Position 2		Position 4		Position 2,4	
	Imidazole	His	Imidazole	His	Imidazole	His
CN	-7.1	-7.0	-6.6	-6.4	-14.0	-14.2
F	-5.1	-4.9	-5.2	-5.4	-10.3	-9.9
Cl	-4.7	-4.6	-4.1	-4.6	-8.7	-8.6
CH2F	-1.6	-1.7	-2.1	-2.4	-3.4	-3.6
OCH3	-1.6	-0.9	-0.9	-2.3	-1.9	-2.3
H	0	0	0	0	0	0
CH3	+1.1	+1.4	+0.7	+1.1	+1.6	+2.4

Table. 3 The amount of pKa shift (ΔpKa) from His (pKa=5.8) and imidazole (pKa=6.9) by the substituted functional groups listed. The amount of pKa shift by a given functional group is very similar in His and imidazole molecules.

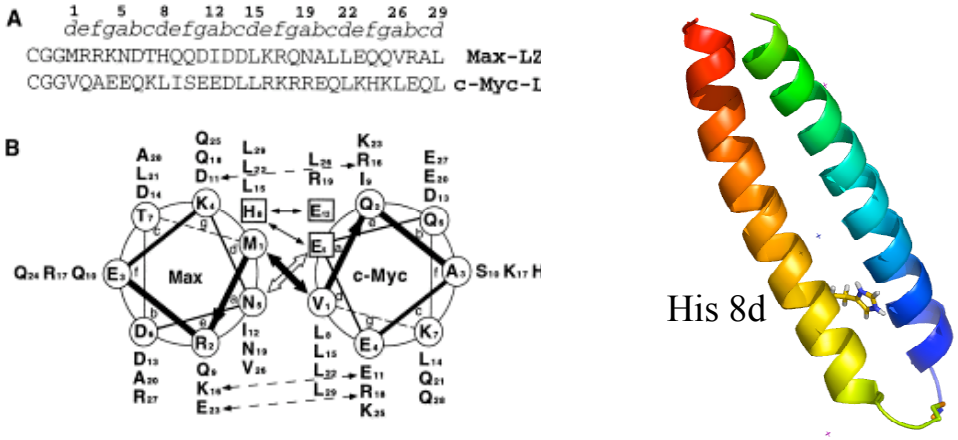


Fig. 4 The structure of the C-Myc-Max heterodimeric leucine zipper. The histidine analog was incorporated to position 8d on the Max helix. This histidine, when protonated, interacts with two glutamate residues on c-Myc through electrostatic interaction. The HCe1 proton (the position 3 as shown in Fig. 3) of Max histidine 8d has a pKa vale of 7.19± 0.03. The C-Myc and Max helices are linked with a disulfide bond in the NMR structure.

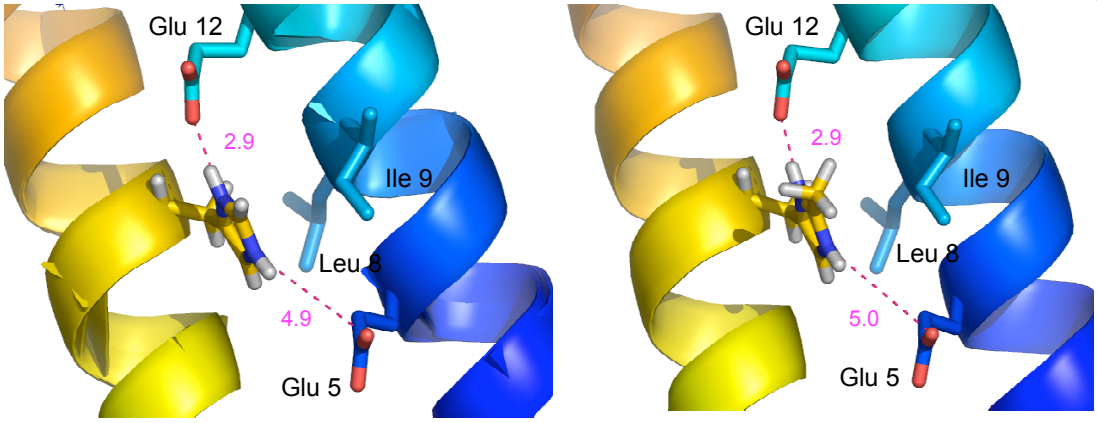


Fig. 5 Comparison of wildtype and 2-methyl histidine binding to the c-Myc helix (in blue). The residues on c-Myc within 5 Å to the histidine imidazole ring are illustrated. The lowest energy rotamer of 2-methyl histidine and the wildtype histidine (in NMR structure) binds very similarly to its surrounding residues on the c-Myc helix.

	Wildtype His charged (kcal/mol)	2CH3-His charged	Wildtype His neutral	2CH3-His neutral	Wildtype His Stabilizing Energy of Protonation	2CH3-His Stabilizing Energy of Protonation	Change in Stabilizing Energy in 2CH3-His compared to wildtype
Glu 5	-21.908	-21.034	0.821	1.245	-22.729	-22.279	0.45
Leu 8	-0.296	-0.381	-1.408	-1.4	1.112	1.109	-0.093
Ile 9	-1.064	-1.608	-2.036	-2.486	0.972	0.878	-0.094
Glu 12	-29.759	-30.152	-5.86	-5.13	-23.899	-25.022	-1.123
Total					-44.544	-45.404	-0.86 **1.9%**

Table 4 The pair-wise interaction energy (in kcal/mol) between residues on c-Myc and His 8d imidazole (on Max). The stabilizing energy of protonation was calculated by subtracting the interacting energy of the neutral His from the charged His. The 2-CH3 substitution caused a small change of 1.9% in the total stabilizing energy of protonation compared to the wildtype His. The highlighted column pairs show that 2-CH3 and wildtype His have similar effects in binding to C-Myc in both the protonated and un-protonated forms.